DOCUMENT RESUME

ED 229 727 CS 007 093

AUTHOR	Frederiksen, John R.; And Others
TITLE	A Componential Approach to Training Reading Skills. Final Report.
INSTITUTION	Bolt, Beranek and Newman, Inc., Cambridge, Mass.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.; Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
REPORT NO	BBN-5295
PUB DATE	Mar 83
CONTRACT	400-81-0031; N00014-80-C-0058
NOTE	206p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC09 Plus Postage.
DESCRIPTORS	*Componential Analysis; *Computer Assisted Instruction; Context Clues; Information Processing; Learning Theories; Models; Phonics; Reading Games; Reading Processes; *Reading Research; *Reading Skills; Secondary Education; *Sequential Learning; *Transfer of Training
IDENTIFIERS	*Automaticity Training

ABSTRACT

Research was conducted to investigate the interactions among component processes of reading and to determine if a hierarchical training model, in which particular reading components are developed sequentially, is an effective way to build reading skills for a target population. Three game-like microcomputer training systems were constructed, each concerned with a critical skill previously shown to pose difficulties for secondary school students with poor reading skills: (1) perception of multiletter units appearing within words, (2) efficient phonological decoding of orthographic information in words, and (3) use of context frames in accessing and integrating meanings of words read in context. Each system was designed to develop the capacity for automatic performance of a particular component by providing a motivating, game-like environment in which to practice the skill. Subjects completed six computer-administered criterion tasks to evaluate their improvement in the targeted skill areas. Results showed that in all cases, subjects were able to reach levels of performance in the trained skills that equalled or exceeded those of high ability readers. There was also strong evidence for the transfer of acquired skills to other functionally related reading components. For subjects who completed the entire training sequence, there were increases in reading speed in an inference task, with no drop in comprehension, suggesting that improvements in the level of automaticity of multiple skill components of reading can reduce the effort required in reading for comprehension. (FL)

****
*   Reproductions supplied by EDRS are the best that can be made   *
*   from the original document.   *
****

Bolt Beranek and Newman Inc. bbn

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Report No. 5295

ED229727

A Componential Approach to Training Reading Skills

Final Report

John R. Frederiksen, Phyllis A. Weaver, Beth M. Warren, Helen P. Gillotte, Ann S. Rosebery,
Barbara Freeman, and Lorraine Goodman

March 1983

Prepared for:
Office of Naval Research and
National Institute of Education

Approved for public release; distribution unlimited.

CS 007 093

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Final Report	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Componential Approach to Training Reading Skills		5. TYPE OF REPORT & PERIOD COVERED Final Report 10/1/79 through 9/30/82
		6. PERFORMING ORG. REPORT NUMBER BBN Report No. 5295
7. AUTHOR(s) John R. Frederiksen, Phyllis A. Weaver Beth M. Warren, Helen P. Gillotte, Ann S. Rosebery Barbara Freeman, Lorraine Goodman		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0058
9. PERFORMING ORGANIZATION NAME AND ADDRESS Bolt Beranek and Newman Inc. 50 Moulton Street Cambridge, MA 02238		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-154-448
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, VA 22217		12. REPORT DATE March 17, 1983
		13. NUMBER OF PAGES 186
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Additional support for the research reported herein was received from the National Institute of Education under Contract No. NEW-NIE-C-400-81-0030. Work on the development of training systems was also carried out under ONR Contract No. N00014-79-C-5014 to Harvard University.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Reading Skills, Transfer of Skills, Individual differences, Automaticity training, Computer assisted instruction, Information processing theories of reading

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The purpose of this research was to develop and evaluate a set of training systems for developing critical skill components of reading. The critical skills that are the focus of training have been shown in prior research to represent particular sources of processing difficulty for secondary school students who have poor reading skills. While the initial motivation for developing component-specific training systems was a desire to investigate experimentally the interactions among component processes of reading, a second and equally important goal was to determine if a

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

hierarchical training model, in which particular reading components are developed sequentially, is an effective way to build reading skills for this target population.

Three game-like microcomputer training systems were constructed, concerned with three skill components: (1) perception of multiletter units appearing within words (the Speed game), (2) efficient phonological decoding of orthographic information in words (the Racer game), and (3) use of context frames in accessing and integrating meanings of words read in context (the Ski-Jump game). Each training system is designed to develop the capacity for automatic performance of a particular component by providing a motivating, game-like environment in which to practice the targeted skill. Furthermore, the training games were designed to ensure that performance of the designated component is mandatory for successful performance, with the trainee receiving immediate feedback concerning his speed and accuracy of responding. Finally, a battery of six criterion tasks was developed. These computer-administered tasks enabled us to evaluate trainee's improvement on the targeted skills and the transfer of skills acquired to the performance of other components. We hypothesized that transfer would follow the lines of functional interaction among skills, from lower level components to components higher in the skill hierarchy.

The results showed that trainees were in all cases able to reach levels of performance in the trained skills that equalled or exceeded those of high ability readers. There was also strong evidence for transfer of acquired skills to other functionally related reading components. Training on the Speed system resulted in improvements in unit detection performance for multiletter units that were included in training. Similar skill improvements were found for units that were not specifically trained, suggesting that the skill developed involves a general ability to encode orthographic information present within a word. The development of this skill had a significant impact on trainees' accuracy of word and pseudoword decoding, and on the amount of text a subject could encode within a single fixation (the span of apprehension). Training with the Racer system brought improvements in both the speed and accuracy of subjects' word decoding. Transfer of training in word decoding to a pseudoword pronunciation task indicated that subjects had acquired an ability to phonologically decode orthographic patterns of English as well as to efficiently identify words. Training with the Ski-Jump game brought an improvement in subjects' ability to establish and use frame-based activations of concepts in semantic memory. Following training, subjects were capable of rapidly recognizing and judging the semantic appropriateness of masked target words that were related to a context frame, regardless of whether the target was a low or a high probability exemplar of the semantically constrained concept. Ski-Jump training also provided some benefit in the development of perceptual encoding skill, although the Ski-Jump game was not as effective as the Speed game in developing that skill. Ski-Jump training also had some impact on more general criterion measures of reading. Four of seven trainees showed improvement in span of apprehension and a like number in accuracy or RT in an inference task, with no drop in comprehension. No such increases occurred for subjects who did not have the entire series of training exercises. This suggests that improvements in the level of automaticity of multiple skill components of reading can reduce the effort required in reading text for comprehension. An analysis of the patterns of transfer of skills acquired in training is presented and principles for optimizing transfer of training in a skill hierarchy are suggested.

Report No. 5295

A COMPONENTIAL APPROACH TO TRAINING READING SKILLS*

John R. Frederiksen, Phyllis A. Weaver
Beth M. Warren, Helen P. Gillotte, Ann S. Rosebery
Barbara Freeman, Lorraine Goodman

Final Report

March 1983

Final Report, Office of Naval Research
Contract No. N00014-80-C-0058,
Contract Authority Identification Number NR-154-448.

---

* The research described was supported by the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-80-C-0058, and also by the National Institute of Education under Contract No. HEW-NIE-C-400-81-0030.

Reproduction in whole or in part is permitted for any purpose of the United States Government.

Approved for public release; distribution unlimited.

TABLE OF CONTENTS

Page

1. INTRODUCTION 7

1.1 General Theoretical Framework 7
1.2 Focus of Training 10

2. METHODS 19

2.1 Design of Training Systems 19
2.2 Game Descriptions 21
2.2.1 The Perceptual Units Trainer: Speed 21
2.2.2 The Word Decoding Trainer: Racer 31
2.2.3 Training in Use of Context: Ski-Jump 46
2.3 Criterion Measures 63
2.4 Subjects and Experimental Design 74

3. RESULTS OF TRAINING 77

3.1 Overview of Methods of Analysis 77
3.2 Speed Training 78
3.3 Racer Training 103
3.4 Ski-Jump Training 124
3.5 Cumulative Summary of Performance for Subjects who Completed the Training Sequence. 159

4. DISCUSSION 167

4.1 Nature of the Skills Acquired 167
4.2 Principles for Optimizing Transfer of Training in a Skill Hierarchy 176

REFERENCES 183

ABSTRACT

The purpose of this research was to develop and evaluate a set of training systems for developing critical skill components of reading. The critical skills that are the focus of training have been shown in prior research to represent particular sources of processing difficulty for secondary school students who have poor reading skills. While the initial motivation for developing component-specific training systems was a desire to investigate experimentally the interactions among component processes of reading, a second and equally important goal was to determine if a hierarchical training model, in which particular reading components are developed sequentially, is an effective way to build reading skills for this target population.

Three game-like microcomputer training systems were constructed, concerned with three skill components: (1) perception of multiletter units appearing within words (the Speed game), (2) efficient phonological decoding of orthographic information in words (the Racer game), and (3) use of context frames in accessing and integrating meanings of words read in context (the Ski-Jump game). Each training system is designed to develop the capacity for automatic performance of a particular component by providing a motivating, game-like environment in which to practice the targeted skill. Furthermore, the training games were designed to ensure that performance of the designated

component is mandatory for successful performance, with the trainee receiving immediate feedback concerning his speed and accuracy of responding. Finally, a battery of six criterion tasks was developed. These computer-administered tasks enabled us to evaluate trainees' improvement on the targeted skills and the transfer of skill acquired to the performance of other components. We hypothesized that transfer would follow the lines of functional interaction among skills, from lower level components to components higher in the skill hierarchy.

The results showed that trainees were in all cases able to reach levels of performance in the trained skills that equalled or exceeded those of high ability readers. There was also strong evidence for transfer of acquired skills to other functionally related reading components. Training on the Speed system resulted in improvements in unit detection performance for multiletter units that were included in training. Similar skill improvements were found for units that were not specifically trained, suggesting that the skill developed involves a general ability to encode orthographic information present within a word. The development of this skill had a significant impact on trainees' accuracy of word and pseudoword decoding, and on the amount of text a subject could encode within a single fixation (the span of apprehension). Training with the Racer system brought improvements in both the speed and accuracy of subjects' word decoding. Transfer of training in word decoding to a pseudoword pronunciation task indicated that subjects had acquired an ability to phonologically decode orthographic patterns of English as well as to efficiently identify words. Training with the Ski-Jump game brought an improvement in subjects' ability to establish and use frame-based activations of concepts in semantic memory. Following training, subjects' were capable of rapidly recognizing and judging the semantic appropriateness of masked target words that were related to a context frame, regardless of whether the target was a low or a high probability exemplar of the semantically constrained concept. Ski-Jump training also provided some benefit in the development of perceptual encoding skill, although the Ski-Jump game was not as effective as the Speed game in developing that skill. Ski-Jump training also had some impact on more general criterion measures of reading. Four of seven trainees showed improvement in span of apprehension and a like number in accuracy or RT in an inference task in which subjects judged the appropriateness of conjunctions describing relations among sentences in a paragraph. Finally, for subjects who completed the entire training sequence, there were increases in reading speed in the inference task, with no drop in comprehension. No such increases occurred for subjects who did not have the entire series of training exercises. This suggests that improvements in the level of automaticity of multiple skill components of reading can reduce the effort required in reading text for comprehension.

An analysis of the patterns of transfer of skills acquired in training is presented and principles for optimizing transfer of training in a skill hierarchy are suggested.

10

ACKNOWLEDGEMENTS

This work would not have been possible without the sponsorship and encouragement of the Office of Naval Research, and the fruitful suggestions and comments offered by Drs. Marshall Farr and Henry Halff, of that office.

We are grateful to the students, faculty, and administration of Cambridge Rindge and Latin School for their cooperation in the conduct of this study. We offer special thanks to Carol Michelson, John Sennott, and Janet Walther, members of the CRLS faculty.

Finally, we wish to thank Marcia Mobilia for her patience and care in preparing the manuscript.

1. INTRODUCTION

This report covers the evaluation of three microcomputer-based training systems for improving the reading skills of young adult poor readers. Each system is designed to develop the capacity for automatic performance of a particular skill component of reading. The theoretical rationale for building a set of component-specific training systems stems from the results of earlier work in which a componential theory of reading skill was proposed and tested (Frederiksen, 1981).

1.1 General Theoretical Framework

A componential theory of reading identifies a set of functionally defined information processing components which, in interaction with one another, accomplished the more complex task of text comprehension. In such a theory, readers differ in the degree to which components have become automated (cf. Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Automatic processes make few attentional demands and can operate concurrently with other components without degrading their efficiency of operation. In contrast, controlled (nonautomatic) processes make demands on general, shared processing resources; when they must operate concurrently with other processes, performance is degraded. A skilled reader possesses many highly automated components, while a less skilled reader has a smaller number of such components, and those may vary considerably within the population of poorly skilled, young adult readers. Thus, while readers may be reliably classified along a single dimension of "general reading ability", the actual sources of low ability may vary considerably from reader to reader.

In our earlier work, we developed a battery of experimental tasks which provide measures of components critical to the reading process. Differences in skill level among high and low ability readers on such tasks were studied and were found to be reflected primarily in the speed or efficiency with which the tasks were performed, rather than in accuracy of performance on the component specific tasks investigated (see Table 1.1 for a list of the components studied and their correlations with four criterion measures of reading ability). Three components were selected as the ones most suitable for training on the grounds that each had a potentially strong impact on the performance of other, higher-level skill; and each, if performed automatically, would reduce the drain on processing resources otherwise required for text understanding. Two of the skills selected are instrumental in processes of word analysis. These include (1) perception of multiletter units appearing within words, and (2) efficient decoding of orthographic information within words. The third skill deals with processes for activation of concepts in semantic memory on the basis of context frames. This component influences the efficiency with which semantic information can be retrieved and integrated in sentence understanding.

Table 1.1

Validity Coefficients[a]

Component	Reading Time For Context	Nelson-Denny Speed	Nelson-Denny Vocabulary	Nelson-Denny Comprehension
	Criterion Measure			
I. Letter Encoding	.17	-.18	-.31	-.20
II. Perceiving Multiletter Units	.20	-.28	-.30	-.29
III. Decoding	.70	-.48	-.62	-.68
IV. Word Recognition Efficiency	.50	-.17	-.35	-.51
V. Speed in Applying Context	.42	-.03	.00	-.21
VI. Extrapolating a Discourse Representation	-.51	.37	.47	.59
VII. Influence of Topicality of Reference	.23	-.17	-.23	-.34
VIII. Semantic Integration of Antecedents	.41	-.11	.08	.02
Mult. R	.74	.63	.73	.76
F (7, 38)	6.48	3.63	6.08	7.50
Prob.	.000	.000	.000	.000

---

[a] Correlations of .25 or greater are underscored.

The perceptual skill is important because it furnishes a basis for efficient and accurate decoding. Decoding and use of context, in turn, while only moderately correlated with each other, are both strongly associated with deficits in general reading skill (cf. Perfetti & Lesgold, 1977; Perfetti & Roth, 1981) and were found to have interactions with higher order components involved in tracing referential relations (Frederiksen, 1982).

1.2 Focus of Training

Word Analysis skills. On the basis of our prior research (Frederiksen, 1981, 1982), we can offer a profile of the word analysis processing of a highly skilled reader, contrasting it with that of a less skilled reader. The skilled reader has a sizeable vocabulary of sight words as well as effective and automatic procedures for decoding unfamiliar words. Efficient decoding for such a reader appears to begin with the automatic recognition of perceptual units larger than the single letter. These units form the basis for subsequent decoding/lexical access. Decoding is effortless and nondisruptive of concurrent processes involved in sentence understanding. Poor readers, on the other hand, appear to be in triple jeopardy. First, their sight vocabularies are smaller, necessitating active decoding on a larger number of occasions. Second, for them the decoding process is effortful and disruptive of other concurrent processing tasks in sentence understanding. Third, their decoding effort must begin with individual letters and only the most common multiletter units. It is reasonable to assume that decoding from individual letters requires more complex rules (cf. Venezky, 1970) than does decoding from a properly chosen and probably rather large "vocabulary" of multiletter units. Phonological decoding from multiletter units involves more consistent rules for pronunciation, and is therefore more amenable to training for automaticity.

With this profile in mind, we sought, through our first training system, Speed, to foster the development of a wide and specific unit "vocabulary", and to build an ability to distribute perceptual/attentional resources over an entire visual array so as to rapidly effect recognition of embedded multiletter units. Both objectives address one of the most well-established sources of deficit in reading skill and one of the most difficult of those to remedy: the poor reader's tendency to attend only to the beginning and perhaps ending of a word that he cannot recognize on sight and then employ a guessing strategy for "decoding" the rest (Harris & Sipay, 1975). Standard reading practice has focused on the first objective--the development of a specific vocabulary of spelling patterns/phonograms which then serve as a basis by which the learner can distinguish among words having common beginning and ending letters. Our instructional focus, in addition to building a specific unit vocabulary, included an effort to modify the poor reader's distribution of attention in word perception.

These newly acquired perceptual skills, in turn, establish a basis for subsequent training in word decoding, the skill addressed in Racer, our second training system. With an increased knowledge of letter clusters and increased ability to detect them quickly within words, poor readers should have less difficulty in developing automatic decoding skill. Racer is designed to provide a practice environment for developing automaticity in word decoding.

Context Utilization Skills. On the basis of two experiments focussing on readers' ability to use semantic information derived from context in gaining access to meanings of words, we can offer a characterization of the nature of expertise in this skill area. In the first experiment (Frederiksen, 1981), high and low ability readers were asked to pronounce target words that were presented following a sentence frame which constrained the meaning of an initially-missing target item. High and low ability readers alike showed reductions in naming onset latencies, the magnitude of which depended upon the degree of contextual constraint. However, these priming effects were in general larger for the high ability readers. Even more interestingly, there were also clear differences among reader groups in the domains of target words for which context priming effects were manifest. High ability readers showed equal degrees of priming for high and low probability words, while low ability readers' priming was restricted to the high probability items only. An explanation for this finding is that highly skilled readers are able to use the semantic information in a context frame to activate concepts in semantic memory. This activation of concept "nodes" provides activation for the domain of lexical exemplars of the primed concept. In other words, it is the semantic relatedness of target words to a sentence frame that determines the degree of priming, not their frequency.

In a second experiment, we developed a more direct test of the proposal that highly skilled readers differ from less skilled readers in their capacity for directly activating semantic categories associated with items in their internal lexicon (Frederiksen, Warren, & Weaver, 1983). In this study, we measured subjects' accuracy and RT in making semantic appropriateness judgments for initially-missing target words appearing after presentation of a context frame. In this task, not only did subjects have to identify the target words, they had to evaluate and integrate their meanings with the propositions represented in the context sentences. In addition to varying the strength of the priming context sentences as in the previous experiment, we introduced target words that were semantically ambiguous, that is, that had two distinct meanings (e.g., "break"), one of which is dominant (e.g., "smash") and the other subordinate (e.g., "tame"). Context sentences were constructed to constrain each of these two possible readings of the ambiguous target (see Figure 1.1).

Our results confirmed the hypothesized reader skill differences. In the absence of strongly constraining context, all readers showed a "primacy effect": the more dominant meanings of the ambiguous words served as the default meaning assignments for those words in moderately constraining contexts. Thus, subjects made significantly more comprehension errors when the subordinate meaning rather than the dominant meaning was required, and had longer RTs in the subordinate condition as well. The interesting differences between reader groups were the changes in this baseline primacy effect when a more strongly constraining context was provided. These results are summarized in Figure 1.2. In the strongly constraining contexts, highly skilled readers were able to utilize context to prime specific meaning categories, thus overriding the primacy effect observed for ambiguous words in moderately constraining contexts. The highly constraining context frames thus allowed the good readers to gain direct access to the meaning required, regardless of whether it was dominant or subordinate. In contrast, for the less skilled readers, primacy effects were still present under the strongly constraining contexts, indicating that they are less able to use semantic information contained in a context frame to activate relevant semantic categories.

D-Sentence

The delicate machinery in the lunar module had to be carefully mounted so that the impact of the landing wouldn't __________ it.

S-Sentence

After the wild horses have been herded into the corral, each cowhand will pick out the one he wants to __________.

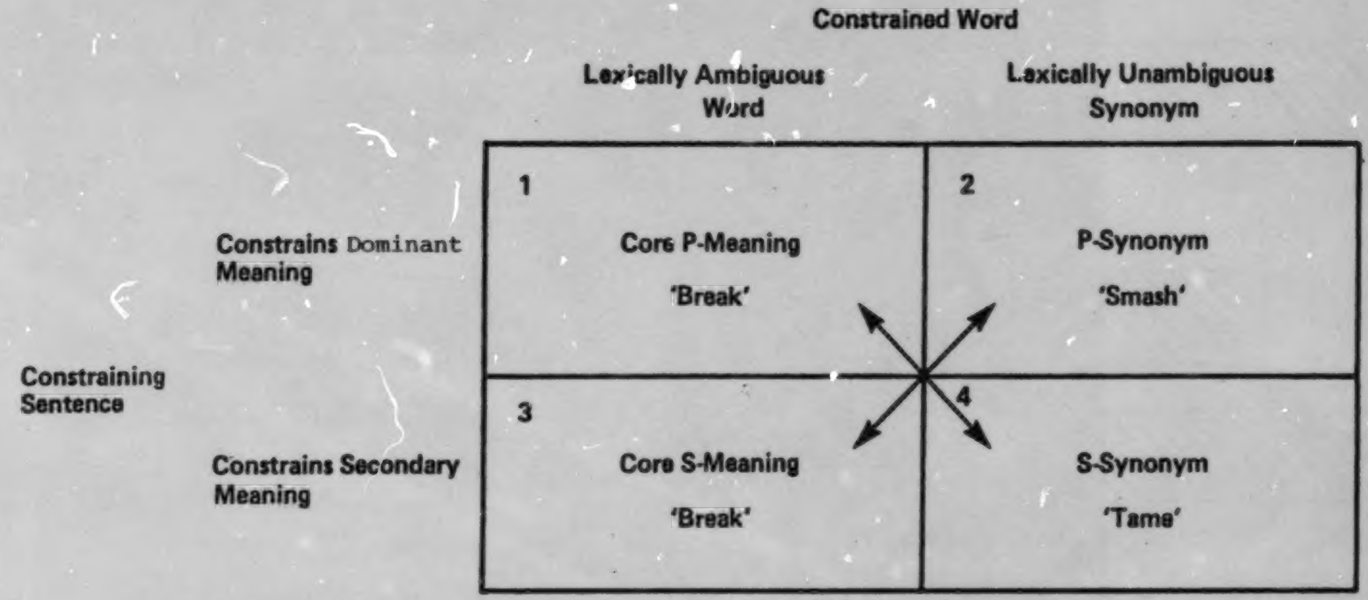

Constrained Word
Lexically Ambiguous Word
Lexically Unambiguous Synonym
Constraining Sentence
Constrains Dominant Meaning
Constrains Secondary Meaning
1
Core P-Meaning
'Break'
2
P-Synonym
'Smash'
3
Core S-Meaning
'Break'
4
S-Synonym
'Tame'

Figure 1.1. Illustrative materials used in the context priming experiment.

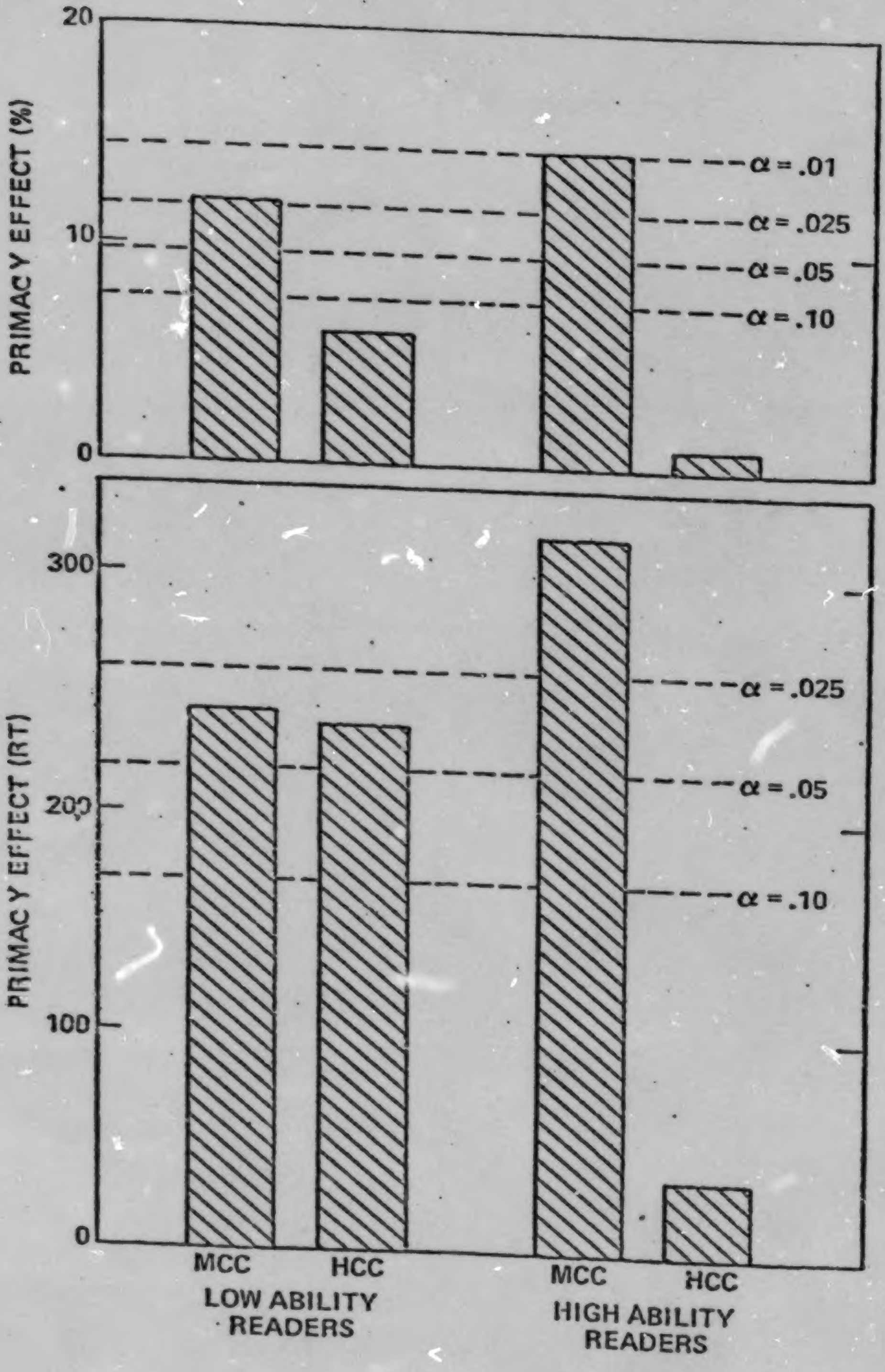


Figure 1.2. Primacy effects obtained in the context priming experiment for moderately constraining contexts and highly constraining contexts. The primacy effect is the difference in performanc in making semantic acceptibility judgements for sentences requiring secondary and dominant meanings of ambiguous target words.

The experimental task that we used to study priming of lexically ambiguous words furnished the basis for our third training system, Ski-Jump. This system provides a practice environment designed to foster the development of an ability to use the semantic information contained in a context frame to activate, in parallel, frame-related concepts in semantic memory. The goal of Ski-Jump training is to develop, again to criteria of automaticity, this capacity to gain access to all words within a domain that are exemplars of frame-related concepts. Such a skill should have an impact on higher level processes of text understanding, including the tracing of referential relations within a text and the analysis of high order relations among propositions in a text (Frederiksen, Weaver & Warren, 1983).

While the initial motivation for developing component-specific instructional systems was a desire to investigate experimentally the interactions among component processes of reading that had been established in studies of intercorrelations among components (Frederiksen, 1982), it is possible that a hierarchical training model in which particular reading components are developed sequentially may be an effective way to build reading skills for a population of subjects who have reached the middle or end of high school without developing adequate reading comprehension ability. By developing three microcomputer games, focussing on the three skill areas described, we sought to test this instructional plan. Our goal is to determine if trainees can develop efficiency in performing the skills required of them in training, and reach levels of performance that are equivalent to those of highly skilled readers who are not given training. We will also evaluate transfer of training from lower level skill components to higher level components. Our hypothesis is that training of lower level components should have an impact on higher level components with which they interact. In particular, perceptual skill training should have an impact on measures of word and pseudoword decoding. Development of efficient decoding and an ability to use context for priming semantically relevant concepts should have an impact on reading efficiency and/or accuracy. And, finally, training of the context priming skill should have an impact on measures of sentence, or paragraph, understanding.

In the Methods section of this report, we present an overview of the design characteristics of our training games, followed by detailed descriptions of each game, as well as specifications for the battery of criterion tasks used to evaluate transfer among skill components. We then present the results of our training studies. Finally, we offer conclusions based on our analyses of these results.

2. METHODS

2.1 Design of Training Systems

The development of automatic skill components requires a practice environment in which several thousand "trials" or instances of skill performance are elicited. Such a heavy burden of practice, in turn, requires that special attention be given to qualities of the training system that have an impact on the trainee's interest and attitude. These requirements led us to the specification of a set of criteria that would govern the essential character of the training tasks, and a list of game-like features that would serve to enhance the overall quality of the practice environment.

Task criteria. For training of automatic skill components to be effective, the performance of a designated component must be *mandatory* for success in the task. Each of our training tasks is therefore designed to focus on developing a particular skill component. Such a practice environment ensures that the target skill, and not some alternative, compensatory skill, will be reliably executed. In addition, the criteria for successful performance on the task --speed and accuracy-- must be directly correlated with performance of the designated component. Successful performance thus involves meeting simultaneous demands of efficiency and accuracy, which are incorporated into the task as direct and immediate forms of feedback. The trainee's immediate goal is to reach and maintain the highest rate of speed that he can manage that is consistent with the accuracy requirements of the task. His longer term goal is to increase that maximum speed without increasing the rate of errors, even as the difficulty of the task is increased.

Game features. To enhance the overall quality of the practice environment and to offset the burden of practice imposed on the trainee, we built some of the clearly compelling features of arcade games into our training tasks.** Some of the features we incorporated are:

- Goals of the game are clearly defined and are represented in visual displays and in scoring rules. The scoring rules are carefully constructed to reflect the costs and rewards associated with various types of actions or responses.
- Feedback is immediate, and speed of responding is as important as appropriateness. The consequences of a successful or unsuccessful action are vividly conveyed through graphic and sound effects.
- The speed of responding is also directly felt in the pace of events within the game.
- The level of challenge in the game is adaptive, with conditions of practice and goals determined by the trainee's current level of expertise.
- Challenge operates as well through the selection of

** See Malone (1981) for a study of intrinsically motivating instructional environments. He identifies challenge, fantasy, and curiosity as the most important features.

training materials which are ordered in difficulty. Trainees are aware of transitions from easier to more difficult materials.

- An element of fantasy, such as a racing motif, is embodied in each game.
- The games involve competition with the computer.
- Performance records are maintained and made available to the trainee.

The specific means by which these criteria and features were realized differed from game to game, as will be apparent in the descriptions of individual games that follow.

2.2 Game Descriptions

2.2.1 The Perceptual Units Trainer: Speed

The first training task, implemented in a microcomputer*** - based game called SPEED, is one in which trainees are required to detect a target multiletter unit when it appears within stimulus words presented in rapid succession. On each training _run_, a student is presented a target unit and then a series of stimulus words in which the unit is either present or absent. His task is to indicate for each word whether or not the target unit is

---

*** The microcomputer employed was an Exidy Sorcerer with 32k of RAM, expansion bus, and running the CP/M system. Programs are coded in CBASIC, with several Z-80 machine-language routines for timing stimuli and responses and for carrying out block transfers within memory. The video monitor used was a Leedex 100.

present by pressing the appropriate response key. The positions of target units within words vary. Rapid detection of units thus requires the trainee to distribute his attention over the entire visual array.

The general goal of the game is for the student to accelerate his rate or speed of unit detections without sacrificing accuracy. The speed at which the stimulus words are presented begins at an initial speed of, for example, 60 words per minute and, depending upon the trainee's performance, increases in the direction of a goal speed which is set 50 units above the initial speed (e.g., 110 wpm). Increases in speed are contingent upon correct detection of the target multiletter unit. Thus, consistently correct detections will lead to a smooth increase in speed until the goal is reached. Attainment of the goal speed, however, does not demand error-free performance. A few errors (which are registered on a set of error lights) are tolerated, but at the expense of speed reductions. Thus, whenever the trainee makes an error, the speed is reduced, permitting him to slow down until he can reestablish accuracy. Accurate detection at the downwardly adjusted speed, in turn, has the effect of again increasing the speed in the direction of the goal. Alternation between errors and correct detections will permit no progress in speed. In order to progress towards the goal speed, the frequency of errors must be kept low.

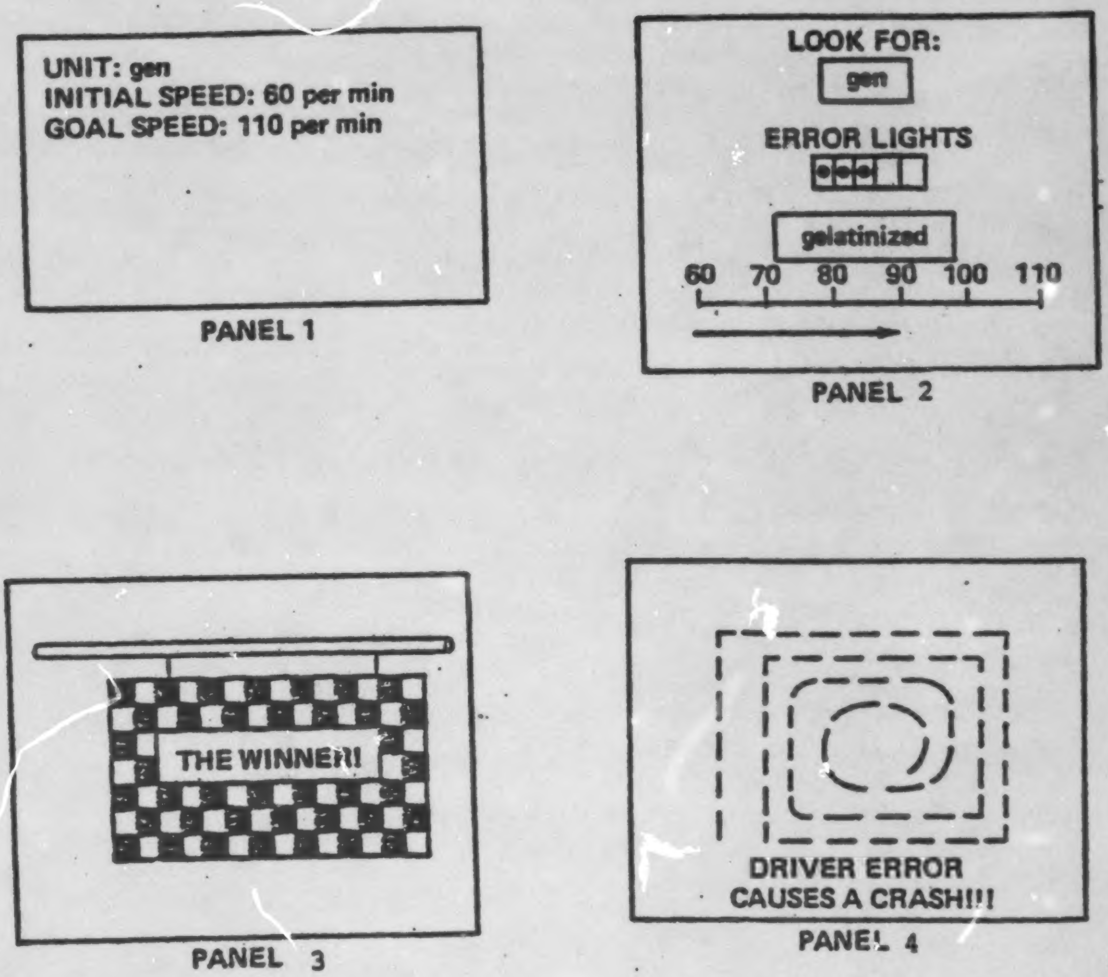
A second means for controlling subjects' errors in detecting units seeks to keep runs of errors to a minimum. Each time an error is made, an error light (one of five such lights) is turned on. A maximum of five error lights is permitted at any one time, and when an error occurs while five lights are on, the game ends with a "crash". The subjects are not limited to five errors total, however. Each correct response causes one error light to be extinguished. The number of error lights on at any one time, therefore, alerts the subject to the need to pay attention to accuracy and tells him how imminent is his potential demise.

Display formats. Initially, the program identifies the unit to be trained, its initial speed, and goal speed as shown in panel 1 of Figure 2.1. Once the game begins, the display is that of panel 2 and contains the following features:

1. the target unit is identified at the top with the heading: "Look for: gen;"
2. five error lights are located directly below the identified target unit;
3. the target words appear in a window located in the center of the display; and,
4. a speedometer is positioned below the display window.

When a student wins or loses the game, one of two special displays appears (panels 3 and 4).

Events within a run. The dynamics of the game are represented in Figure 2.2 in which changes in the rate of word


UNIT: gen
INITIAL SPEED: 60 per min
GOAL SPEED: 110 per min
LOOK FOR:
gen
ERROR LIGHTS
gelatinized
60
70
80
90
100
110
THE WINNER!!
DRIVER ERROR
CAUSES A CRASH!!!

PANEL 1

PANEL 2

PANEL 3

PANEL 4

Figure 2.1. Display formats in the game SPEED.

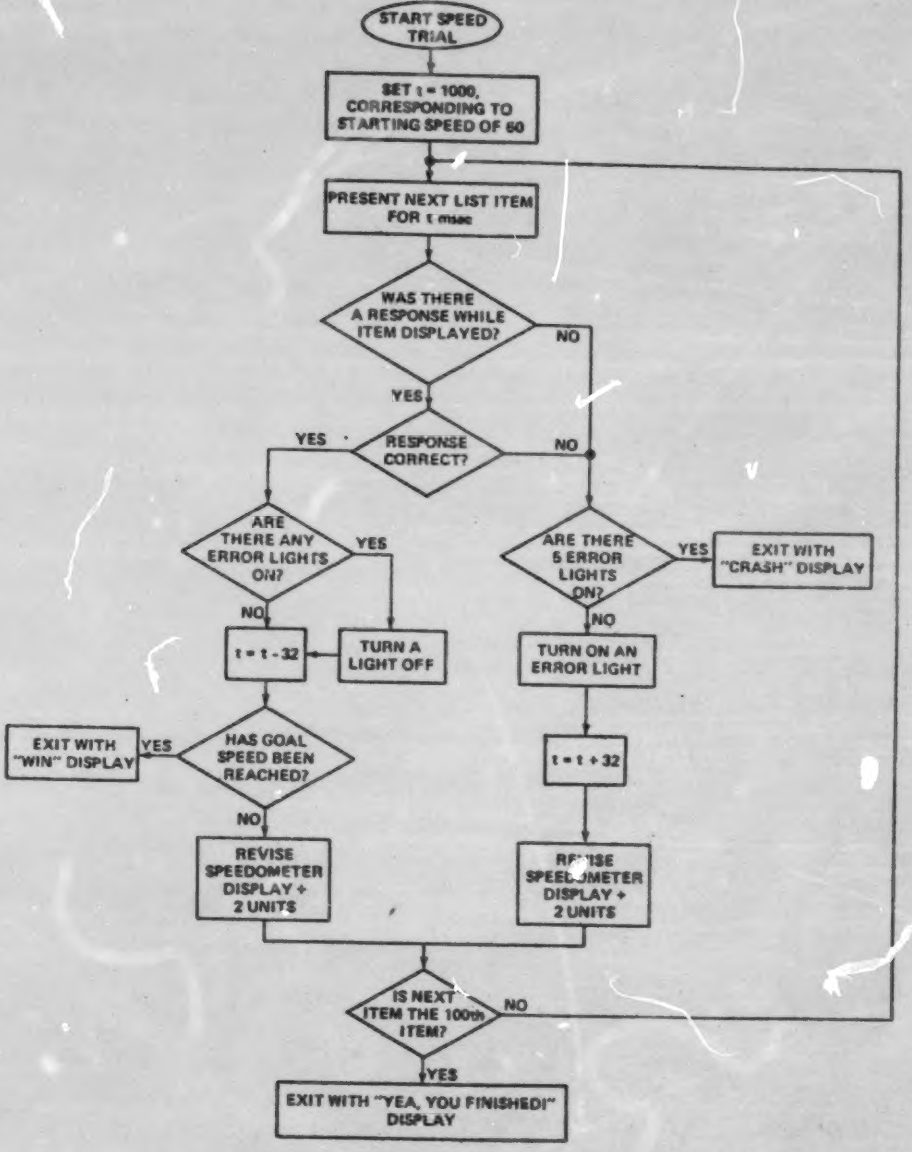

START SPEED TRIAL
SET t = 1000, CORRESPONDING TO STARTING SPEED OF 60
PRESENT NEXT LIST ITEM FOR t msec
WAS THERE A RESPONSE WHILE ITEM DISPLAYED?
NO
YES
RESPONSE CORRECT?
YES
NO
ARE THERE ANY ERROR LIGHTS ON?
YES
NO
t = t - 32
TURN A LIGHT OFF
ARE THERE 5 ERROR LIGHTS ON?
YES
EXIT WITH "CRASH" DISPLAY
NO
TURN ON AN ERROR LIGHT
t = t + 32
EXIT WITH "WIN" DISPLAY
YES
HAS GOAL SPEED BEEN REACHED?
NO
REVISE SPEEDOMETER DISPLAY + 2 UNITS
REVISE SPEEDOMETER DISPLAY + 2 UNITS
IS NEXT ITEM THE 100th ITEM?
NO
YES
EXIT WITH "YEA, YOU FINISHED!" DISPLAY

Figure 2.2. Flow chart representing the dynamics in the game SPEED.

presentation and in the error lights and speedometer are related to the student's responses on each stimulus word presented within a run. The speed of item presentation is set by varying the time each item remains displayed. A starting speed of 60 corresponds to an initial display time of t = 1000 msec. Trainees must respond within that interval or the trial is counted as an error. If the response is correct, two things happen: first, if there are any error lights currently on, one is turned off; second, the time t used for the next stimulus presentation is reduced by 32 msec (corresponding to an increment of 2 speed units on the speedometer). If this upwardly adjusted speed matches the set goal speed, the trainee wins the game and is shown the "WIN" display. In the event the goal speed has not been reached, the speedometer display is revised, and if items remain in the list, the next item is presented.

When the trainee's response to a new item is incorrect (or when he fails to respond in the allotted time), the speedometer and error lights undergo other adjustments. If five error lights are already on, the trainee "CRASHES". Otherwise, an additional error light is registered (panel 2, Figure 2.1) and the time for the next item is increased by 32 msec, with the corresponding 2-unit decrement registered on the speedometer.

Whenever the program attempts to present the next list item, it checks to see if the list has been exhausted. If it has, the message, "Yea, you finished!" appears in the stimulus window.

Subsequent runs. Initial and goal speeds on the second and subsequent runs on a unit are governed by the trainee's past levels of performance on that particular unit. At the beginning of a run, the program accesses the final performance level on the previous run and sets a new initial speed 30 units below that. For example, if a trainee reached a final speed of 110, the new initial speed would be set at 80 and the corresponding goal speed set at 130.

Materials. Sixty units were included for training in the Speed system. An additional 20 units served as controls (see Table 2.1). Selection of these 80 units was based on the results of a computer simulation (PARSYL; see Weaver, Frederiksen, Warren, Gillotte, Freeman and Goodman, 1982 for further details) which parsed words into multiletter perceptual units, and then applied a set of heuristics for joining the units into pronounceable syllables or vocalic units. PARSYL enabled us through an iterative process to converge on the smallest unit vocabulary that would at the same time permit accurate and efficient decoding.

For each of the 80 units selected, we composed an inspection list of 100 stimulus words, containing the following distribution of word types: 50 test words that contain the target unit; 25 fillers that are highly similar to the test words in terms of unit resemblance ("gen" changed to "gem", "ger", "pen", etc.),

Table 2.1

Units Used in the Speed Training System*

1	un	an	ly	re
2	ac	be	th	ab
3	ism	ing	cl	ure
4	tion	ist	ence	ish
5	ty	ter	ous	ound
6	ble	ple	ite	di
7	age	ace	pre	ful
8	ate	sp	con	ent
9	br	ph	for	ness
10	pl	ash	gen	bit
11	pro	ight	min	ize
12	com	eck	por	cal
13	ick	ord	ast	ance
14	and	ile	tr	ven
15	ire	ex	ler	im
16	sh	ob	bi	il
17	mis	ake	as	sion
18	der	col	sive	ale
19	less	sen	sin	wh
20	de	ver	ock	ant

*These four columns represent the 80 units (60 trained, 20 control) used in the Speed system. Units were assigned to one of four groups of 20 units each. The groups were, in general, matched on unit length, positional likelihood, and frequency of occurrence. Five units in each group were designated as control units.

word length, and configuration; and another 25 fillers that are dissimilar to the test words except in word length. In addition, the distribution of unit positions within the inspection list for a given unit was controlled to reflect positional likelihoods in a more extended corpus of words. Words varied in length, from 4 to 18 letters. Table 2.2 contains sample inspection lists for the units cl and gen.

Sequencing of units. An active training group (ATG) of 6-8 units was maintained at any given stage in training. The ATG included a sampling of both easy and hard units, with units classified for each subject on the basis of a difficulty ranking derived from their pretests on the unit detection task. In that task, a unit was presented, followed by a series of stimulus words, some of which contained the target unit. For each subject, mean RTs over those stimulus words were calculated for each unit and then ranked in difficulty.

The units in the ATG were presented in a random round robin fashion until performance on a unit reached a criterion speed. When the criterion speed was reached for any unit, that unit was deleted from the ATG and replaced with another unit of comparable difficulty. As training progressed, however, subjects showed such marked improvements in the rates at which they acquired the units, regardless of their difficulty level, that it became unnecessary to replace a mastered unit with one that was

Table 2.2

Sample Inspection Lists for the Units cl and gen

Unit: cl		Unit: gen	
Test Word	Filler Word	Test Word	Filler Word
1. claimed	chaired	1. generality	gynecology
2. clanged	freeway	2. generating	treasonous
3. clarion	chalice	3. generation	goniometer
4. classed	grenade	4. generators	grenadiers
5. clauses	slavish	5. generosity	redundance
6. cleanly	predate	6. generously	transplant
7. cleanup	slaver	7. generic	gnostic
8. cloture	diapers	8. genesis	probity
9. clamping	clovenly	9. genetic	gnocchi
10. classics	freehold	10. genteel	squalid
11. clearest	slattern	11. genuine	germane
12. clemency	frequent	12. detergency	wholesaler
13. climaxes	alembic	13. indulgence	outrageous
14. clinical	prejudge	14. insurgency	surpassing
15. clocking	slipknot	15. negligence	pentethion
16. closed	presto	16. pathogenic	orthogonal
17. closet	elicit	17. geneology	resultant
18. clouds	trepan	18. genocidal	geologize
19. clover	eleven	19. gentility	character
20. clucks	prefix	20. genuinely	gerundive
21. clumsy	slalon	21. generally	invisible
22. cloak	greed	22. generals	geminate
23. claws	slams	23. generous	pellagra
24. clasp	green	24. genitive	ganister
25. click	slack	25. genotype	gnathite
26. cliff.	wrest	26. collagen	plethora
27. clamorous	elsewhere	27. divergent	packaging
28. clarifies	presidium	28. negligent	intrigant
29. clattered	challenge	29. resurgent	obstinacy
30. clientele	prejudice	30. gender	geisha
31. clippings	sluggishly	31. genial	gnawed
32. acclimate	treasurer	32. gentry	ponder
33. declaring	foolhardy	33. generically	reincarnate
34. inclusive	frequency	34. gentlemanly	gelatinized
35. reclining	yielding	35. eugenic	mercury
36. cyclist	kindred	36. regency	lexical
37. enclave	evelong	37. emergent	integers
38. enclose	lamprey	38. ontogeny	micawber
39. unclean	gosling	39. legendary	pageantry
40. declared	entirety	40. pungent	longing
41. exclaims	mealtime	41. exigency	quotient
42. included	fortress	42. endogenous	caliginous
43. unclench	choleric	43. octogenarian	hallucinated
44. clarifying	figurehead	44. heterogeneity	permanganate
45. classifies	albuminous	45. interagency	spherometer
46. clodhopper	discretion	46. generatively	octosyllable
47. clad	elan	47. regenerate	manzanella
48. clef	urea	48. tangential	loggerhead
49. exclusions	amsliorate	49. cryogenic	inorganic
50. unclasping	discrepant	50. ontogenesis	sympetalous

comparable in difficulty. Instead, a new unit was simply selected randomly from the remaining pool.

Mastery criteria for units were developed for individual subjects, and were adjusted upwards as a subject's skill at the game increased. Mastery criteria were set initially at the maximum speeds a subject reached in 4-6 runs with the initial set of units. They were then adjusted upwards in 10-20 wpm increments when the initial criterion speeds were reached by the subject in 2-3 runs. Similarly, the starting speeds for newly introduced units were increased as the subject gained in skill. Final mastery criteria varied from 150 wpm to 210 wpm.

2.2.2 The Word Decoding Trainer: Racer

This training task requires the subject to rapidly and accurately pronounce a set of words as they appear one at a time in the cells of a matrix displayed on the computer monitor. The game consists of two parts, Racer and Sound Trap, which in combination serve to train both speed and accuracy in decoding words. A major difficulty in designing a computer-based system for training decoding is the limited speech recognition capability of the microcomputer, and our consequent inability to judge directly the correctness of the subjects' pronunciations of stimulus words. In our system, we met this difficulty by requiring the subject to recognize correct pronunciations of stimulus words in a "Sound Trap" test following completion of the

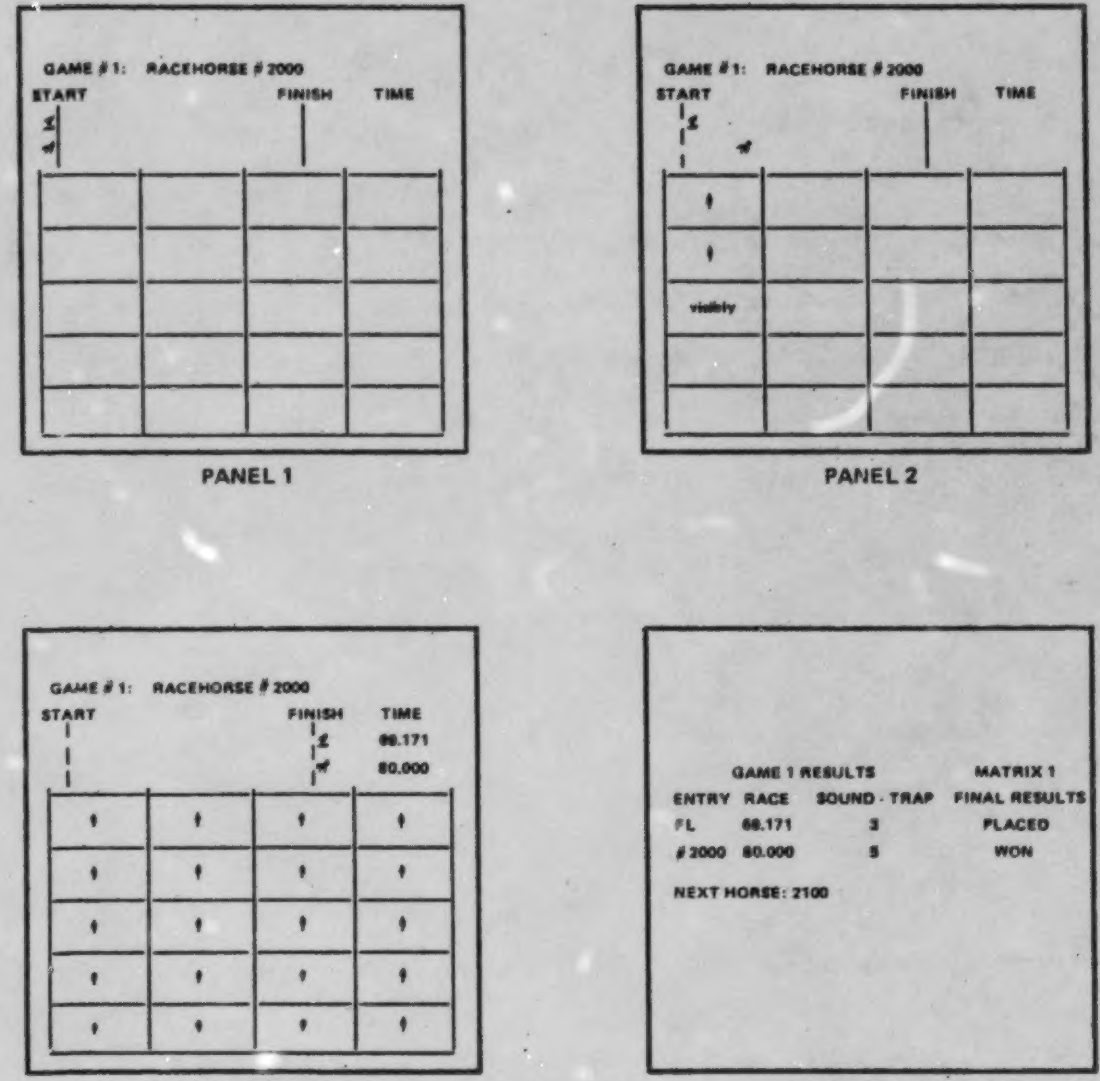
"Racer" portion of the game. During the Race, we use a Cognivox speech interface[****] simply as a voice key to measure speed of responding. We can in that way measure the trainee's speed of decoding and maintain a proper degree of time pressure on him. In the "Sound Trap" that follows, the subject listens to a series of word pairs (delivered in the current implementation using an audio cassette player) and is required to select which of the two words of a pair appeared during the Race. This procedure has proven very effective in motivating accuracy in decoding. While few people can actually recall the words from the Race matrix, subjects can easily recognize the words they have pronounced. Correct choices in Sound Trap can therefore be reliably made, provided only that the display words in the Race have been accurately pronounced.

The goal of Racer is for the trainee to decode a series of twenty words as quickly and accurately as he can. The race involves a competition between a sailboat, representing the trainee, and a horse, running at a rate based on the trainee's own earlier performance. The horse sets the current level of challenge for the trainee who is impelled by the advancing horse to pronounce each word rapidly. Whenever a word appears, the horse starts to move and continues to move at a fixed pace until the trainee pronounces the word. The sailboat moves a fixed distance for each word -- 1/20th the distance to the finish line -- but only when the student makes his pronunciation. If the student responds quickly, he can move his sailboat ahead of the horse and go on to win the race. Immediately following the race, the cassette machine automatically begins playing the Sound Trap in which a subset of the 20 display words is presented, each paired with a similar-sounding foil (for example, "catch" and "cash"). The trainee's task here is to listen to each pair and judge which of the words appeared in the race matrix. Since he does not see the word-pairs, he must rely totally on the sound and on his recognition memory for the words he decoded during the race.

[****] The Cognivox is a voice input-output device that plugs into the Sorcerer microcomputer.

The trainee's superordinate goal is to run against swifter horses in succeeding races. In order to earn the right to run against faster horses and thus to progress in the game, the trainee must win the race _and_ make no more than one error out of eight Sound Trap pairs. More errors will force him to run against increasingly slower horses. Thus, speed and accuracy jointly determine the trainee's progress in the game.

Display formats. The opening display in Racer (panel 1, Figure 2.3) identifies the current game number, and the number (or speed) of the horse against which the trainee will compete. Also shown is the display matrix containing 20 cells. Once the


GAME # 1: RACEHORSE # 2000
START
FINISH
TIME
visibly
68.171
80.000
GAME 1 RESULTS
MATRIX 1
ENTRY
RACE
SOUND - TRAP
FINAL RESULTS
FL
68.171
3
PLACED
# 2000
80.000
5
WON
NEXT HORSE: 2100

PANEL 1

PANEL 2

PANEL 3

PANEL 4

Figure 2.3. Representative displays employed in the Racer game.

race begins, the display looks as it does in panel 2, Figure 2.3 showing:

1. the relative positions of the horse and sailboat as they progress towards the FINISH line;
2. the number of pronunciations attempted thus far (in this case, two); and
3. the next word in the series to be pronounced (e.g., visibly).

When the race ends (panel 3, Figure 2.3), the time in seconds for the horse and sailboat each to cross the finish line are displayed.

The display for Sound Trap is shown in Panel 4, Figure 2.3, and includes:

1. the racing times for the just concluded race for both the trainee and horse;
2. a tabulation of the trainee's performance on the word pairs tested in Sound Trap, with correct responses noted on the trainee's line and incorrect responses credited on the horse's line;
3. a number indicating the speed of the next horse (which changes if the subject makes more than one error);
4. the final results representing combined results on Racer and Sound Trap.

In addition, when the trainee makes an error on a word pair, the correct word is displayed in the area adjacent to the "next horse" number.

Events within a run. The dynamics of Racer and Sound Trap are depicted in Figures 2.4 and 2.5, respectively.

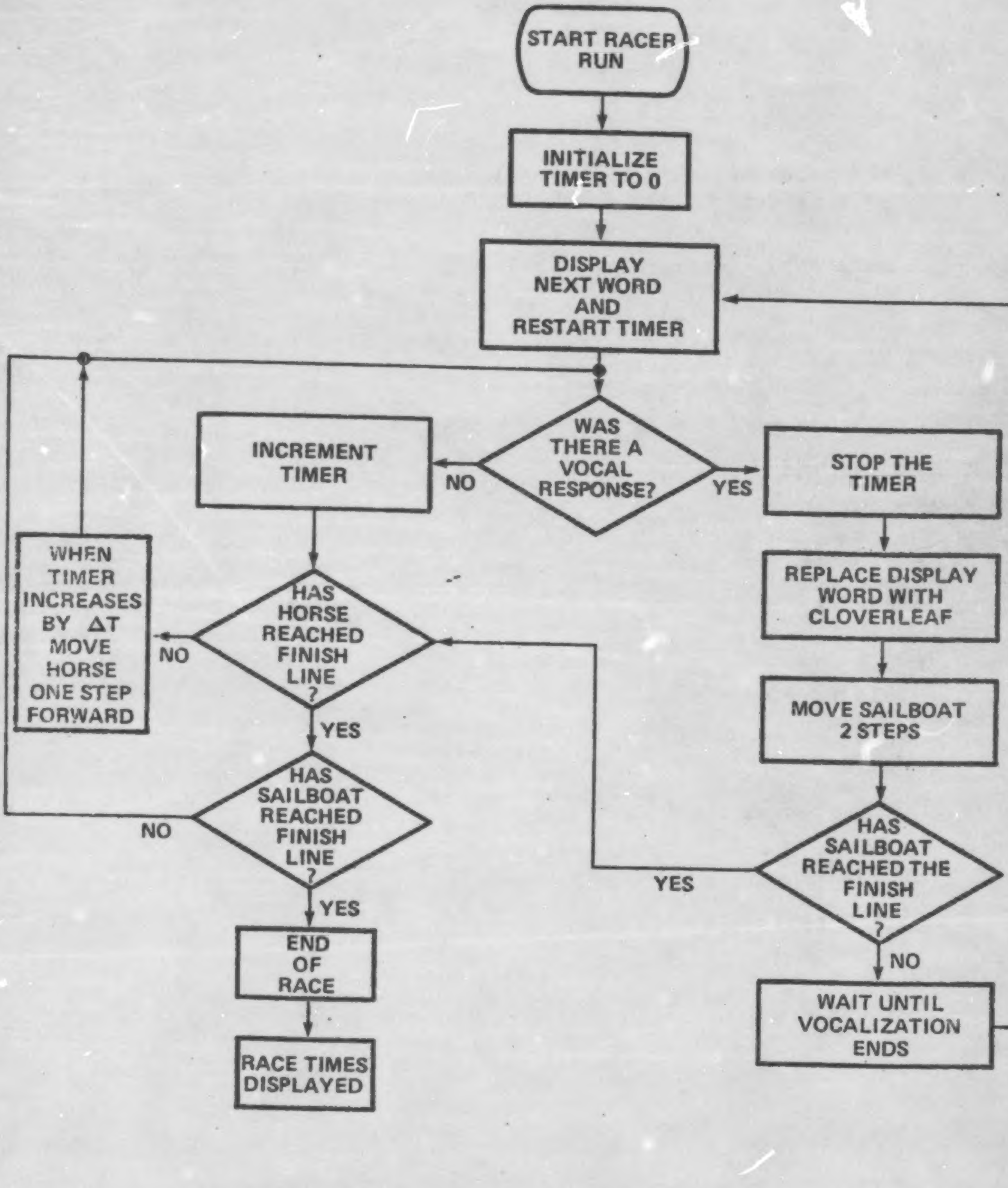

START RACER RUN
INITIALIZE TIMER TO 0
DISPLAY NEXT WORD AND RESTART TIMER
WAS THERE A VOCAL RESPONSE?
NO
YES
INCREMENT TIMER
STOP THE TIMER
REPLACE DISPLAY WORD WITH CLOVERLEAF
WHEN TIMER INCREASES BY ΔT MOVE HORSE ONE STEP FORWARD
NO
HAS HORSE REACHED FINISH LINE ?
YES
MOVE SAILBOAT 2 STEPS
HAS SAILBOAT REACHED FINISH LINE ?
NO
YES
HAS SAILBOAT REACHED THE FINISH LINE ?
YES
NO
END OF RACE
WAIT UNTIL VOCALIZATION ENDS
RACE TIMES DISPLAYED

Figure 2.4. Flow chart representing the dynamics in Racer.

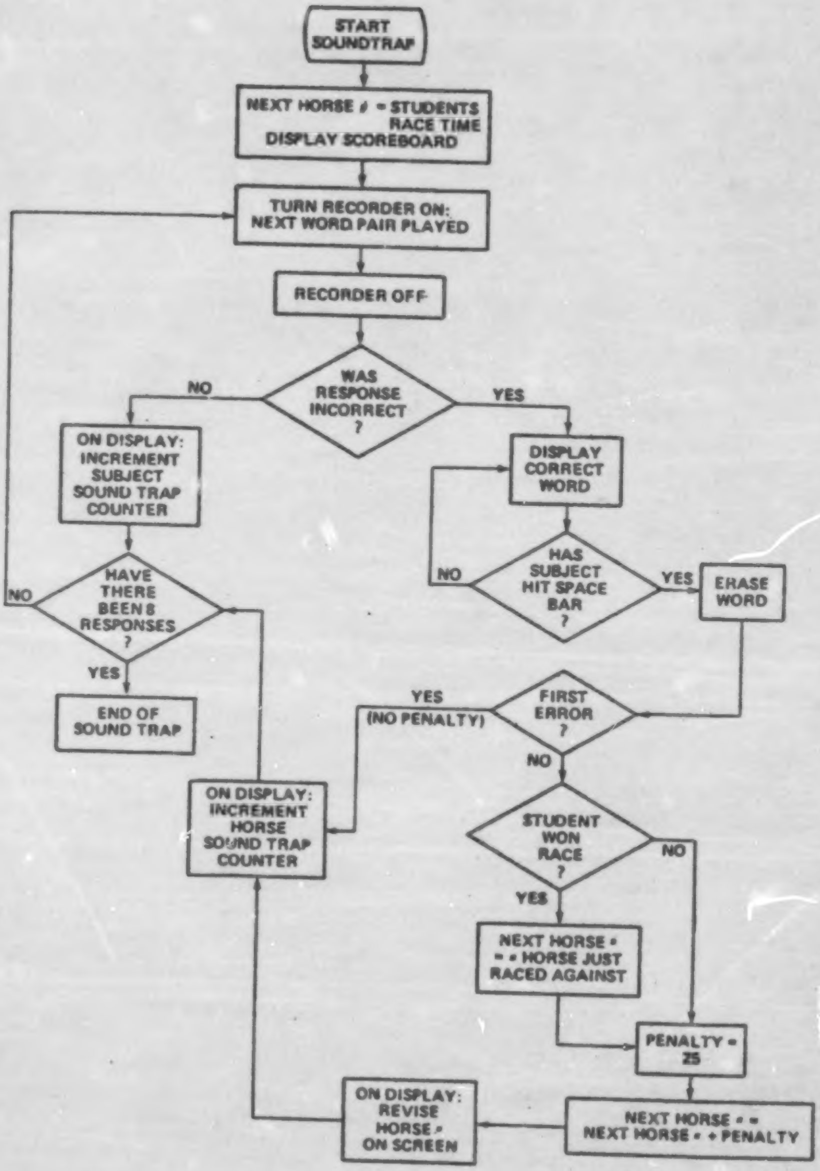

START SOUNDTRAP
NEXT HORSE # = STUDENTS RACE TIME DISPLAY SCOREBOARD
TURN RECORDER ON: NEXT WORD PAIR PLAYED
RECORDER OFF
WAS RESPONSE INCORRECT ?
NO
YES
ON DISPLAY: INCREMENT SUBJECT SOUND TRAP COUNTER
DISPLAY CORRECT WORD
HAVE THERE BEEN 8 RESPONSES ?
NO
YES
END OF SOUND TRAP
HAS SUBJECT HIT SPACE BAR ?
NO
YES
ERASE WORD
FIRST ERROR ?
YES (NO PENALTY)
NO
ON DISPLAY: INCREMENT HORSE SOUND TRAP COUNTER
STUDENT WON RACE ?
YES
NO
NEXT HORSE # = # HORSE JUST RACED AGAINST
PENALTY = 25
ON DISPLAY: REVISE HORSE # ON SCREEN
NEXT HORSE # = NEXT HORSE # + PENALTY

Figure 2.5. Flow chart representing the dynamics in Sound Trap.

In Figure 2.4, the progress of the race between the sailboat (or trainee) and the horse is shown to be related to the occurrence of the trainee's vocal response to each of the 20 display words. At the moment the first word is displayed, a clock having a total of 40 ticks begins to count, and continues to do so until a vocal response is registered. The clock then stops ticking and only resumes when the next stimulus word is displayed. Each tick is equal in milliseconds to the number assigned to the horse. (If, for example, the horse has been assigned #1000, the clock ticks every second. And, two times the horse number gives the time allowed per word if the subject and the horse were to run a tie.) For each tick of the clock, the horse advances 1 step. From the starting gate to the finish line there are 40 steps. When a vocal response is registered, the display word is replaced with a cloverleaf, and the sailboat is advanced 2 steps, or 1/20th of the distance. Thus, if the vocal response takes less than 2 ticks, the sailboat will gain on the horse. The computer then waits for the subject's vocal response to end, and after an interval that could vary from 200 to 350 msec depending on the subject*****, the next word is displayed and the clock restarted. If, on the other hand, the sailboat has

---

***** The delay interval following vocalization was introduced in order to ensure that a memory "trace" would be established for each word pronounced. The interval was initially set at 350 msec, but subjects were allowed to shorten the interval to as low as 200 msec if they wanted to speed up the pace of the game.

reached the finish line (that is, 20 words have been displayed and pronunciations attempted) ahead of the horse, then the horse automatically advances the remaining steps to the finish and the race ends. The competitors' times are then displayed.

When a vocal response to the display word is delayed, the clock keeps ticking and the horse advances steadily until a response is made. Although the delay may actually be long enough to permit the horse to cross the finish line, the game nonetheless continues until the trainee has attempted pronunciations for all of the 20 stimulus words and has moved his sailboat across the finish line. Racing times are then displayed.

The race results are then transferred to the Sound Trap scoreboard. Figure 2.5 illustrates the course of events in Sound Trap, focussing on the scoring rules relating accuracy in Sound Trap to the speed of the horse in the next run of Racer. Initially, the next horse in the competition is equal to the trainee's race time, regardless of whether he crossed the finish line first. Thus, having the possibility of racing against increasingly faster horses depends entirely on the speed with which the trainee makes his way through the matrix. Penalties arising from inaccurate performance in Sound Trap, however, can result in his running against a horse slower than the one he earned the right to run against in the race just completed. These dynamics are described next.

Referring again to Figure 2.5, Sound Trap begins automatically with the playback of the first of eight cassette-recorded word pairs. The trainee is required to select which of the two words in the pair appeared in the display matrix by pressing the appropriate response key. If he answers correctly, he is credited with one point, and the next word pair is played. If, on the other hand, he answers incorrectly, the horse is credited one point, and the correct word is displayed on the screen until the trainee presses the space bar to continue. No additional penalties are associated with the first error. On the second error, the student loses Sound Trap, and the penalties that are applied guarantee that in the next race he will run against a slower horse. These penalties vary depending on whether or not the trainee finished the race ahead of the horse. If the trainee has won the race, then the "next horse" number is first set back to the number of the horse he just raced against (it being the slower of the two times for the race), and then a further 25 points are added to that number. If the trainee has lost the race, the "next horse" number is already slower than the horse just raced against, so the 25 point penalty alone is added to that number. Losing Sound Trap thus ensures that the subject will run against a slower horse, regardless of whether he has won or lost the race. Further 25 point penalties are added for each additional error in Sound Trap. This rather stringent set of scoring rules evolved in the first several weeks of training with

Racer. Prior to their introduction, it was possible for a student to lose Sound Trap, yet go on to race against a faster horse in the next race. Thus, errors in Sound Trap were not costly enough, resulting in an imbalance favoring speed of decoding over accuracy.

Sound Trap ends when responses to all eight of the word pairs have been made. The trainee may then elect to go on to the next race.

Materials and sequencing of materials. The general principle governing the construction of Racer matrices was to build a sequence of matrices that gradually increase in difficulty. The difficulty of a matrix was initially manipulated by varying the length of words it contained, where length was measured by the number of syllables in a word. A second variable we planned to manipulate was the frequency class of the words. A Racer dictionary of 10,530 words was prepared, and random access CP/M disk files created for each of the syllable lengths and frequency classes described in Table 2.3. A CBASIC program was developed to construct the Racer matrices. The program reads a "specification" formula which indicates the number of syllables and frequency classes of words to be selected in creating the Racer matrices. It then creates a sequence of Racer matrices as output, each of which contains a random selection of words from the dictionary conforming to the specifications. Each time a

Table 2.3

Racer Dictionary

Number of Syllables	Frequency Class	SFI Range	Number of Items
1	High	47.0-50.0	500
2	High	47.8-50.0	500
2	Moderate	37.6-40.1	1,000
3	Moderate	33.5-40.4	2,000
4	Moderate	35.1-47.1	1,250
5	Moderate	34.6-50.7	515
2	Low	24.6-30.4	1,000
3	Low	24.6-33.5	2,000
4	Low	24.6-35.0	1,250
5	Low	24.6-34.5	515
-	--	24.6-50.0	10,530

dictionary word is chosen, a marker is placed on that entry, and no word is eligible for re-selection until 5 intervening matrices are built. Sets of 15 matrices were created for each prospective training session, since it was estimated that this number of runs could be completed in a thirty minute training session. The composition of the Racer matrices is given in Table 2.4. The standard matrices developed using the Racer dictionary are those from A1 through E3. They contain words drawn from the high and moderate frequency categories.

Item difficulty within each matrix of the standard series was systematically varied. Within each column of a matrix, item difficulty (measured by syllable length) decreased as subjects progressed from the top to the bottom row. Likewise, average length of words within a column decreased from left to right across columns.

Two of our subjects had difficulty in developing efficient decoding skills when trained using the standard matrices. To meet this need, a special series of "consistent" matrices (the Z-series) was created. The rationale was that training focussing on a single decoding rule or principle would enable subjects to practice those rules and develop automaticity in applying them. When these subjects reached acceptable speeds on the consistent matrices, they would then be moved into the standard series. Table 2.5 shows the composition of each of the Z-series matrices

Table 2.4

Summary of Racer Matrices

Name	Word Types
Z1	1 syllable, short vowels
Z2	1 syllable, long vowels
Z3	2 syllable, short vowels
Z4	2 syllable, long vowels
Z5	initial blends, 1 & 2 syll., short vowels
Z6	initial blends, 1 & 2 syll., long vowels
A1	1 syllable, mixed vowels
A2	16 one-syll., 4 two-syll., mixed vowels
A3	14 one-syll., 6 two-syll., mixed vowels
A4	All two-syllable, mixed vowels*
B1	All two syllables
B2	16 two-syllables, 4 three-syllables
B3	5 one-syll., 9 two-syll., 6 three-syll.
E1	8 two-syll., 10 three-syll., 2 four-syll.
E2	5 two-syll., 12 three-syll., 3 four-syll.
E3	3 two-syll., 11 three-syll., 6 four-syll.

*All of the remaining matrix types had mixed vowels.

Table 2.5

Sample Z-Series Matrices

Matrix Cell	1 syll., short vowel	1 syll., long vowel (includes digraph vowels, markers)	2 syll., short vowel in stressed syllable	2 syll., long vowel in stressed syllable	initial blends & consonant digraphs, 1&2 syll., short vowel in stressed syllable	initial blends & consonant digraphs, 1&2 syll., long vowel in stressed syllable
1	selves	time	catcher	riding	knock	sleet
2	tap	seem	penguin	going	wrecker	stroke
3	mud	wave	after	eating	shrill	grating
4	fox	may	visit	wisest	blemish	write
5	jest	tone	rocket	hiding	stuff	stride
6	kept	paint	basket	needy	chigger	throne
7	list	roll	emptied	maintained	shot	cream
8	felt	bees	sanded	cables	strand	bridle
9	went	both	wicked	giant	classic	blue
10	tell	pail	backing	taping	snap	statement
11	had	lays	welfare	people	chest	gray
12	bat	iced	pollen	payment	fluffy	shame
13	pens	jays	camel	gamely	tramp	whine
14	rat	safe	helping	safest	strip	brine
15	rock	roped	saddest	behind	drag	sprain
16	fist	face	indoor	keepers	clan	child
17	held	wait	wettest	legal	chock	stolen
18	add	lakes	inner	needed	trend	trite
19	pups	weed	seven	miners	prim	stipend
20	cam	old	fixer	seated	sledding	flame

used in training, and a sample matrix of each type. Z-series matrices were not drawn from the standard dictionary, but were built from special files containing words exemplifying the decoding rules. Subjects were not told in advance the composition of the matrices they were receiving, but discovered for themselves the principle involved.

The level of difficulty at which each subject could comfortably begin intensive practice was established in the first few training sessions by having all subjects practice initially on the easiest matrix level (the A-level). Subjects for whom the A-level proved challenging were continued on that level, while those who progressed rapidly through it were advanced immediately to the B-level. Thereafter, subjects generally completed 30 matrices (600 words) at each difficulty level. Over this number of matrices, our subjects typically showed substantial improvements in decoding efficiency. They maintained these improvements as each new level of difficulty was introduced and, by the end of practice at each, showed still greater improvements in speed of decoding. In those cases where a subject made slow progress with a particular set of matrices, additional practice at that level was provided until gains were registered.

2.2.3 Training in Use of Context: Ski-Jump

In this task, the trainee is required first to read a sentence which has one of its final words deleted. After reading this context sentence, he is presented a series of target words, one at a time, and for each word he must judge whether it is semantically appropriate to the context. Each target word is displayed on the screen in several flashes, the duration of each successive flash increased by a constant of 18 msec. Each target word is presented with pre- and post-exposure masks, and is followed by an inter-stimulus interval which represents the amount of time between exposures that a student has to integrate contextual and perceptual information in making his judgment. The goal in this game is to judge the appropriateness of the word on the earliest flash possible. The earlier the student responds, the higher will be his score. In order to be able to identify the word when it is barely visible, the student must make use of the context that has been provided. Thus, to succeed at the game the student must learn to integrate rapidly the clues he derives from context with the bits of visual information he picks up when the target word is flashed.

A beginner at this game will typically require many flashes before he sees a word and can only improve his performance when the context sentence is highly constraining, providing very obvious clues about the identity of the word. After practicing for a while, however, he learns to profit from context clues that are less obvious.

Each run of the Ski-Jump game involves a team of skiers, representing the target words for that sentence, setting out on a slope from which each is to jump. The length of a jump is determined by how early the student judges the acceptability of a target word by pressing the appropriate response key. Early recognition leads to a long jump, later recognition to a shorter jump. The final score for the run is based upon an average of the total number of feet jumped by the team of skiers. In addition, there is an incentive for the player to be accurate as well as early in judging the target word. Each error is costly in that it counts as a jump of zero feet.

As in Speed and Racer, this game also increases in challenge as the trainee progresses in skill. Initially, the intervals between exposures of the target words are fairly long (2000 msec). Later in training, the durations between exposures are considerably shorter (10 msec). As a result, high scores come to depend more and more on the successful and speedy integration of contextual clues, however obvious or subtle, with the minimal visual information derived from very brief exposures of the target words.

Display Formats. Initially, the program displays a sentence context in which an underscore has been substituted for the deleted target word as shown in Figure 2.6. Once the game begins, the display is similar to that shown in Figure 2.7 and contains the following features:

He knew it would be a major

challenge to present his ideas

to this ______________________.

Figure 2.6. Sample Sentence Context Used in the Ski-Jump System.

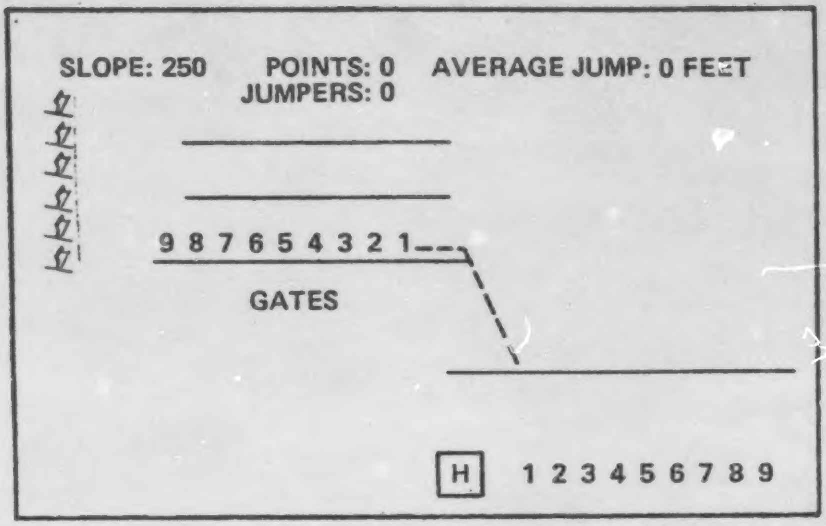

SLOPE: 250
POINTS: 0
AVERAGE JUMP: 0 FEET
JUMPERS: 0
9 8 7 6 5 4 3 2 1
GATES
H
1 2 3 4 5 6 7 8 9

Figure 2.7. Display format at the start of a run in Ski-Jump.

1. The current slope (representing the duration of the inter-stimulus interval) is identified in the top left corner.

2. The indicators for point count, jumpers and average jump are to the right of the slope number (these are set to zero at the beginning of a run).

3. A number of skiers are aligned in a column down the left side of the display, each skier standing for a target word.

4. A viewing window is represented by two horizontal lines positioned to the right of the skiers. During a "play", the target word appears here.

5. Nine gates, represented by the numbers 9-1, are arranged along the jumping platform, directly below the viewing window.

6. The landing area is below and to the right of the jumping platform, connected to it by a dotted line which represents the hill.

7. Nine bins, positioned directly beneath the landing area indicate at what exposure a response was made. The skiers collect in these bins over the course of a run, forming a histogram of the subject's performance.

8. The "hospital" is located directly to the left of the bins beneath the jumping platform.

At the end of each run the student is informed of the outcome of his performance by one of three displays illustrated in Figure 2.8.

The final display allows the student to review his performance, as shown in Figure 2.9. It contains the following features:

1. The sentence context

2. All target words aligned in one of two columns:

GO ON TO SLOPE XXX

or

STAY ON SLOPE XXX

or

GO BACK TO SLOPE XXX

Figure 2.8. Three displays illustrating the possible outcomes of a Ski-Jump run.

```
      Press space bar to continue....

               He knew it would be a major

               challenge to present his ideas

               to this ______________________.

YES:  * gathering                NO:     elbow

      * club                           * doormat

        meeting                        * mountain

      * group                          * basketball

      * audience                       * bookcase
```

Figure 2.9. Final display format in Ski-Jump.

  a) a "yes" column indicating those words which fit the context, and

  b) a "no" column indicating foils.

3. An asterisk that appears to the left of each correctly judged word.

Events within a run. The dynamics of Ski-Jump are represented in Figure 2.10 in which the interplay between a student's performance on each word within a run and the scoring system which determines the level of difficulty for the subsequent run are illustrated. Within one run of Ski-Jump, a student reads a sentence of either high or low constraint in which one of the final words has been deleted. He is then presented with a series of target words, one at a time. His task is to judge as quickly as possible and within the fewest number of exposures whether the target word is semantically appropriate to the prior sentence context.

The trainee begins a run by pressing the space bar which initiates display of the sentence context. A second press of the space bar clears the screen and calls up the Ski-Jump display (see Figure 2.7). A third press of the space bar releases the first skier and a target word is presented in a succession of very brief exposures. Each exposure of the target word is preceded by a premask of 20 msec and followed by a postmask of 32 msec.

The initial exposure duration of a target word is 18 msec

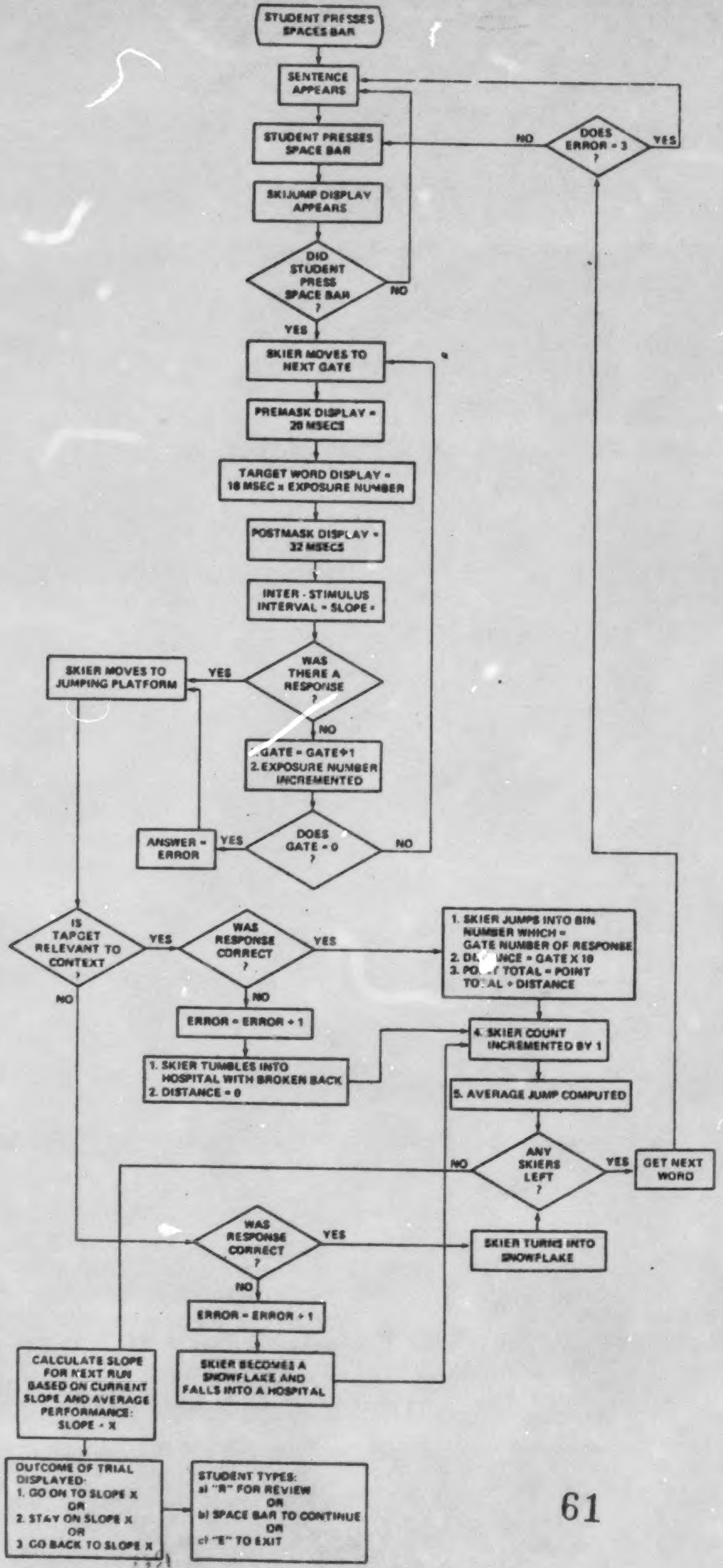

STUDENT PRESSES SPACES BAR
SENTENCE APPEARS
STUDENT PRESSES SPACE BAR
SKIJUMP DISPLAY APPEARS
DID STUDENT PRESS SPACE BAR ?
NO
YES
SKIER MOVES TO NEXT GATE
PREMASK DISPLAY = 20 MSECS
TARGET WORD DISPLAY = 18 MSEC x EXPOSURE NUMBER
POSTMASK DISPLAY = 32 MSECS
INTER - STIMULUS INTERVAL = SLOPE =
WAS THERE A RESPONSE ?
YES
SKIER MOVES TO JUMPING PLATFORM
NO
1. GATE = GATE+1
2. EXPOSURE NUMBER INCREMENTED
DOES GATE = 0 ?
YES
ANSWER = ERROR
NO
DOES ERROR = 3 ?
NO
YES
IS TARGET RELEVANT TO CONTEXT ?
YES
WAS RESPONSE CORRECT ?
YES
1. SKIER JUMPS INTO BIN NUMBER WHICH = GATE NUMBER OF RESPONSE
2. DISTANCE = GATE X 10
3. POINT TOTAL = POINT TOTAL + DISTANCE
NO
ERROR = ERROR + 1
1. SKIER TUMBLES INTO HOSPITAL WITH BROKEN BACK
2. DISTANCE = 0
4. SKIER COUNT INCREMENTED BY 1
5. AVERAGE JUMP COMPUTED
ANY SKIERS LEFT ?
YES
GET NEXT WORD
NO
NO
WAS RESPONSE CORRECT ?
YES
SKIER TURNS INTO SNOWFLAKE
NO
ERROR = ERROR + 1
SKIER BECOMES A SNOWFLAKE AND FALLS INTO A HOSPITAL
CALCULATE SLOPE FOR NEXT RUN BASED ON CURRENT SLOPE AND AVERAGE PERFORMANCE: SLOPE = X
OUTCOME OF TRIAL DISPLAYED:
1. GO ON TO SLOPE X
OR
2. STAY ON SLOPE X
OR
3 GO BACK TO SLOPE X
STUDENT TYPES:
a) "R" FOR REVIEW
OR
b) SPACE BAR TO CONTINUE
OR
c) "E" TO EXIT

Figure 2.10. Flow chart representing the dynamics in Ski-Jump.

and on each successive exposure, the duration is incremented by 18 msec. (Specifically, the exposure duration for Exposure #1 = 18 msec, Exposure #2 = 36 msec, Exposure #3 = 54 msec, etc.) Thus, the target words become clearer and clearer to perceive with each successive exposure. A word is flashed until the student responds by pressing a designated key or until a maximum of nine exposures have been displayed.

The duration of the inter-stimulus interval (ISI) which separates the postmask of one exposure from the premask of the next determines the level of difficulty within a run, and is given by the slope. The slope (ISI) is fixed for a given run. For example, if the slope is currently set at 250, then for every word in the run, the trainee has 250 msec in which to indicate his judgment of the contextual appropriateness of that target word before the next stimulus exposure occurs. For each word, the number of exposures required before the subject responds is recorded. (No response or an erroneous response count as 10 "exposures".) The mean number of exposures required is calculated and used to determine the slope on the next run. If the mean number of exposures is less than or equal to 4, the slope is reduced; if it is greater than 6, the slope is increased, and if it is between 4 and 6 or is equal to 6, the slope remains the same on the next run. The subjects overall goal is to reach the fastest slope possible (a slope of 7) by increasing his speed of responding, and thereby decreasing the average number of exposures of the stimulus word he requires to make a correct judgment of semantic appropriateness.

When a student responds to a target word, one of four things can happen depending upon whether the target is a contextually relevant word or a foil, and on whether the response is accurate or inaccurate.

Contextually relevant words. When a trainee's response to a contextually relevant word is _correct_, the skier (representing a target word) moves down the hill and into the bin whose number corresponds to the gate at which the student responded. The distance earned equals the gate number multiplied by 10, and this value is added to the point count, which is the cumulative sum of all the distance counts in the run. For example, a correct response at Gate 8 to the first target word earns a distance of 80 feet and a correct response at Gate 7 to the subsequent word earns a distance of 70 feet. Together, these jumps add 150 to a trainee's point total. Thus, the distance score informs the student about his performance on an individual word within a run and the point score records his cumulative performance in a run. A second indicator of a trainee's ongoing performance is his _average jump score_. The average jump is calculated by dividing the total number of points earned by the number of words to which a student has responded. It is recomputed with each new response and is used to inform the trainee what the final outcome of a run is likely to be. For instance, the average jump of two correct responses at Gate 8 and two correct responses at Gate 7 is 75 feet since (2(80)+2(70))/4=75.

On the other hand, if a trainee's response to a relevant word is incorrect or if he fails to respond within the allotted time (ISI) following the 9th flash, the skier plunges into the hospital with a broken back and the trainee is penalized for his error. In this case, the distance jumped equals zero feet. Thus, the average jump is recalculated with an unchanged numerator and a larger denominator yielding a lower average jump. (Referring to our sample run: (2(80)+2(70)+0)/(4+1)=60 feet.)

Foils. When the trainee responds correctly to a contextually irrelevant word, the skier turns into a snowflake and disappears. No points are earned for a correct response to a foil, and no changes are made in the various score counts. Conversely, a trainee is penalized for an inaccurate response to a foil. In this case, the skier is transformed into a snowflake but now the snowflake falls into the hospital and is scored as an error. The error is recorded as a response, and the average jump score is recomputed with a larger divisor (e.g., (2(80)+2(70)+0)/(5+1)=50 feet).

After each response to a target word, the trainee may call up the next word or he may choose to review the sentence context. However, the sentence context is automatically redisplayed on the occurrence of a third error.

When all the target words have been displayed, the run is over and the student is informed of the outcome of his performance by the average jump score he has earned. An average jump greater than or equal to 60 feet earns a more difficult slope, a score from 40 feet to 60 feet maintains the present level of difficulty, and a score of less than 40 feet indicates that the next run will be less challenging. If our sample run had ended after the first four words, the trainee would have earned an average jump of 75 feet, enabling him to move on to a more challenging slope. At the end of each run, the trainee has the option of examining a display containing a review of his prior performance, of continuing on to the next run, or of stopping play.

Subsequent runs. The level of challenge for a subsequent run of Ski-Jump is governed by a trainee's previous performance and is determined by the current slope and the average jump score earned (or, equivalently, the average number of exposures) in the run. The calculation of a new slope occurs whenever the average jump score is outside of the 40 to 60 foot range, as described above. The *new* or *revised slope* to be used on the next run is given by:

$$\text{New Slope} = \text{Old Slope} \times \frac{\text{Average Exposure Number}}{5}$$

The old slope is multiplied by a ratio of the average exposure number divided by 5. For example, if the subject required an average of 3.16 exposures on a run (corresponding to an average jump****** of 68.3 feet) and the current slope was 250, his new slope would be 250x(3.16/5)=158. If, on the other hand, his average exposure number had been 7.8 (corresponding to an average jump of 22 feet), then his new slope would be 250x(7.8/5)=390. The use of a ratio rather than a difference in comparing the subject's average exposure number with the middle value on the Ski-Jump scale (5) ensures that the revised slope will represent a proportionate change from the current slope (within a range from .2 to 2, depending on a subject's performance). This reflects the fact that large absolute changes in slope early in training (when the slopes are large) will have a smaller impact on task difficulty than will the same changes late in training when the slopes are already small. Proportionate reductions (or increases) in slope using factors such as .8 (representing a 20% reduction) or 1.2 (representing a 20% increment) are more likely to have a comparable impact on task difficulty at different stages in training.

Materials and sequencing of materials. The materials were drawn from a corpus of constraining context sentences and words developed in an earlier study (Frederiksen, 1978). The context

---

****** The average jump is related to the average exposure number by the equation:
Average Jump = 100-10X Average Exposure Number.

sentences for Ski-Jump training consisted of 150 low constraining sentences and an equal number of moderate-to-high constraining sentences. The degree of constraint was scaled using a group of 15 high school subjects who were given the context sentences and asked "to fill in the missing word". For each sentence, the number of different words that were produced by our subjects was calculated. The number of word types produced for high constraining sentences ranged from 1 to 14 words, with a mean of 6.6, while that for low constraining sentences ranged from 7 to 23 words, with a mean of 13.2 words.

For each of the context sentences, a set of semantically appropriate target words was collected. The target words varied in their probability of occurrence for a given context, where probability was defined as the frequency with which they were generated by the same group of 15 high school subjects used in scaling the context sentences. The number of target words for each sentence in our corpus varied from 6 to 8, with 4 to 6 semantically appropriate and the remainder inappropriate ("foils"). For each context sentence, the appropriate words were presented in the order low probability preceding high probability. Foils were randomly intermixed. Table 2.6 contains a sample of the materials used in Ski-Jump training.

For training purposes, the sentences were grouped by degree of constraint into a number of separate files, each of which

Table 2.6

Sample Materials Used in Ski-Jump System

High Constraining Context:

He later confessed that he had not, after all, enjoyed the ______________________.

Appropriate Target Words	Inappropriate Target Words
tour	coiffeur
exhibit	heron
concert	
play	
show	
movie	

Low Constraining Context:

The teacher labored through the night but by dawn every last exam had been ______________________.

Appropriate Target Words	Inappropriate Target Words
finished	bethrothed
graded	boiled
marked	
checked	
corrected	

Note. Appropriate target words are listed in the order low to high probability.

contained from 8 to 22 sentences, an average of 15 sentences per file. The presentation order of sentences within a file was randomized for each subject. Trainees were generally able to complete one such file per session. Subjects began training with the highly constraining context sentences. When they achieved mastery criteria for these sentences, they were moved to the more subtle, low constraining sentences. The mastery criteria required subjects to reach the fastest slope possible in the game (Slope 7) while at the same time responding accurately within the first 3 target word exposures, and to maintain that level of performance over one session of 15 sentences. Training was terminated when subjects met the same criteria for the less constraining sentences.

2.3 Criterion Measures

A series of criterion tasks was administered to our trainees before and after training on each system. Measures derived from these tasks allowed us to evaluate the performance gains on the particular target skills we were training and the transfer of those gains to other more representative reading tasks. Four of the criterion tasks were designed to evaluate component skills targeted by particular training games. These are the unit detection task (Speed), the pseudoword and word pronunciation tasks (Racer), and the context priming task (Ski-Jump). Two more representative reading tasks include the span of apprehension task and the inference task. A measure of reading rate was also derived from performance on the inference task. Finally, the Nelson-Denny Reading Test was readministered to several of our subjects at the end of their participation in the instructional experiment.

1. Unit detection task. The first criterion task provides a measure directly related to the objective of Speed training. A target unit is first presented, followed by an inspection list of 32 words, 20 of which contain the target unit and the remaining 12 of which are distractors, divided equally into a similar set (containing sets of letters similar to the target unit) and a dissimilar set (containing no similar letter groups). The distribution of unit positions within the inspection list is controlled to reflect positional likelihoods in a more extended corpus of words, with the exception that no units appear at the ends of words. Words vary in length ranging from 6 to 9 letters. At a viewing distance of 22 inches, each letter space represents 18 minutes of visual angle. Each target unit is presented on the Leedex 100 Video Monitor for 150 msec, and is followed by a visual mask of equal duration. Subjects respond to each stimulus word by pressing one of two response keys, indicating whether the target unit was present or absent. Reaction times for unit detections are measured by the Sorcerer microcomputer using a Z-80 machine language timing loop to an accuracy of 6 msec. Correctness of the subject's response is also recorded.

The test covered a total of 32 units, 16 of which were from the set trained in the Speed system and 16 of which were control units not included in the Speed game. Units were restricted in length to 2 and 3 letters. Since in this task individual RTs were measured for each detection trial, it was possible to analyze mean RT as a function of unit length, unit position within a word, and unit set (trained vs. control units).

2. Pseudoword pronunciation task. The second criterion task is a pseudoword pronunciation task (Frederiksen, 1978) in which subjects pronounce pseudoword items that were derived from actual English words by changing a single vowel or vowel digraph (e.g., BRENCH, derived from BRANCH). The set of pseudowords includes 19 orthographic forms, and these include variations in length (4, 5, or 6 letters), number of syllables (1 or 2), and type of vowel (simple vowel, vowel digraph, silent -e marker). For each form, 16 pseudowords beginning with each of 8 phonemes were constructed from either a high or low frequency base word. There are a total of 19 x 8 x 2=304 stimulus items. These items were divided into four sets of 76 items, which were matched on initial phoneme type and frequency, resulting in four equivalent forms of the pseudoword pronunciation task.

Stimuli are presented on the screen of the video monitor, and subtend visual angles ranging from 1^{o} 14' (4-letter items) to 1^{o} 50' (6-letter items) at a viewing distance of 22", which is the approximate distance for our subjects. Stimuli are presented for a duration of 200 msec. Subject's vocal reaction times are measured from the onset of the stimulus to the onset of vocalization, using a Cognivox speech interface connected to the Sorcerer microcomputer. Reaction times are measured by the microcomputer using a machine language timing loop; individual RTs are timed to an accuracy of 6 msec. The accuracy of each response is also rated by the experimenter. Results are summarized using two statistics: mean RT for correct pronunciations and percentage of correct responses.

3. Word pronunciation task. The third criterion task is similar to the pseudoword task, except for the substitution of words for pseudowords. The words substituted were matched to the pseudowords they replaced in orthographic form, initial phoneme, and frequency class. However, they were not the base words from which the set of pseudowords was derived.

4. Span of apprehension task. Our third criterion task permits us to measure the span of apprehension--the width of the effective visual field--for a target phrase presented alone or as the completion of a context phrase. As shown in Figure 2.11, our subjects view a series of displays on the screen of the Leedex Monitor, each of which is made up of 3 frames. Frame 1 contains a context paragraph. Subjects read the paragraph at their own rate. When they reach the end, they fixate a spot appearing in

Frame 1

> They notice that the heat changes from hour to hour. So the day is carefully planned. They know it is hottest during the afternoon. So they do not work then. Instead, they rest. They may take a nap. As

Frame 2

Frame 3

> a rule they do their jobs later.

Figure 2.11. The sequence of displays used in the context condition. Frame 1 contains the context and is subject terminated. Frame 2, presented for 200 msec, contains the fixation point. Frame 3 contains the test phrase, and is also presented for 200 msec.

the final line of the display and press a response key. Frame 2 is then presented for 200 msec followed by the test phrase in frame 3, also for 200 msec. The subject's task is to report as many words or word fragments as he can see in the test line. The subject's response is typed by the experimenter, and transferred into a Sorcerer CP/M disk file. The response measure is the distance in letter spaces from the left-most to the right-most correct letter of the phrase. The subject's vocal onset RT is also measured and recorded.

In each session, subjects are presented a total of 40 test phrases for which a context passage is presented prior to the test exposure, and 40 test phrases in which the prior context passage is omitted. Alternatively, blocks of 8 context and 8 no context trials are presented. The passages used in the experimental task are taken from the Degrees of Reading Power Test (State of New York, Board of Regents, 1976, Forms X and Y). The test passages cover five levels of text readability, chosen to represent approximately equal steps on the Bormuth scale. For each readability level, there are two sets of eight consecutive test passages that together make up a mini-essay on some topic of general interest. One of these mini-essays is assigned to the context condition, and the other to the no-context condition.

Our measure of the subject's ability to encode words within a single fixation is the mean visual span for test phrases presented without accompanying context. Our measure of the subject's ability to utilize context is the visual span when a prior context is provided.

5. Context priming task. The context priming task is derived from the Ski-Jump training task, by setting the parameters of the instructional system to certain fixed values. As in the Ski-Jump system a context sentence is first presented. When the subject presses a response key, the Ski-Jump display is drawn on the screen and the successive exposures of the first target word begin. Each exposure of the target is preceded and followed by a masking stimulus, presented for 20 msec and 32 msec respectively. The initial exposure is 18 msec (1 "paint" of the CRT), and is incremented by that amount on each succeeding exposure. The subject's task is to judge whether the target word is semantically appropriate to the context in which it occurs, and respond by pressing the appropriate key on the keyboard. Following a trial, the subject is given the option of viewing again the context sentence, or proceeding to the next test word.

The materials were drawn from a corpus of words and constraining sentences developed in an earlier study (Frederiksen, 1978). For each sentence, a set of words was collected, all of which fit the semantic domain created by the context. The number of test words varied from 6 to 8 with 4 to 6 semantically appropriate and the remainder inappropriate ("foils"). The appropriate target words varied in their probability of occurrence in the context, where probability was defined as the frequency with which they were generated by a set of 15 high school subjects who were given the context sentence and asked to "fill in the missing word". The appropriate words were presented in the order low frequency preceding high frequency. Foils were randomly intermixed.

Two versions of the context priming task were created, each containing 12 low constraining sentences and a similar number of moderate-to-high constraining sentences. The degree of constraint of a context sentence was scaled using the same group of 15 high school subjects used in scaling response probability. For each sentence, the number of words (types) falling within the semantic domain was calculated (i.e., the number of different words that were semantically consistent with the context). High constraining contexts within our corpus ranged from 1 to 14 words with a mean of 6.6 words, while low constraining contexts ranged from 7 to 23 words, with a mean of 13.2 words. In the criterion task, the high and low constraining sentences were alternated.

Dependent variables derived from the context-priming task include: the mean RT for foils, the mean RT for high probability target words (occurring last in the set of test words for a given sentence), and mean RT for low probability targets occurring as the first test item (the one immediately following the context sentence). RT included time from the onset of the first exposure of the stimulus[*******] until the subject's response. In addition, the mean percent of correct judgments of semantic acceptability was recorded for each class of stimuli. These dependent variables were measured for both high and low constraining contexts.

6. Inference task. A test of inferential comprehension was constructed from materials used in an earlier experiment (Frederiksen, Weaver and Warren, 1981; 1983). The general purpose of the task was to test subjects' ability to infer high order relations among sentences that are not explicitly marked in a text, and to demonstrate their understanding of such a relation by selecting from a test pair of conjunctions the one representing the relation. Their correctness in judging the appropriate conjunction and their RT in making their judgment constitute the two principal dependent variables in the analysis.

The experimental task is illustrated in Figure 2.12. The subject is presented a series of three-sentence passages, one sentence at a time. The first two sentences contain propositions that together lead to or imply the proposition contained in a third sentence. The third sentence begins with a blank space

---

[*******] RTs were actually measured from the onset of the premask which occurred a fixed 20 msec prior to the stimulus.

[Subject presses space bar to see first sentence.]

1. A flourishing cash crop of shellfish used to be a major source of revenue for states of the eastern seaboard.

[Subject presses space bar to see second sentence.]

2. In recent years, industrial wastes and oil spills have changed the environmental conditions along the coast so that shellfish no longer thrive.

[Subject presses space bar to see third sentence with missing conjunctions

3. ______________, the marketing of shellfish is no longer so lucrative.

[Subject presses space bar to see choice of 2 conjunctions. The subject should be prepared to answer with keys marked 1 or 2 as quickly as possible.]

4. 1. As a result

   2. Despite this

[After the subject answers, the incorrect answer will be erased, leaving only the correct answer on the display for 1/2 second. Then the fifth display will follow.]

5. A major source of revenue for eastern seaboard stated used to be:
   1. oil 2. shellfish 3. tourism

[Subject gives his answer by pressing the keys marked 1, 2, or 3. If he is correct, the next story will be presented. For an incorrect answer, the correct answer is displayed on the screen, and the subject is directe to press the space bar to go on to the next story.

There are 20 stories per group.

Figure 2.12. The sequence of displays used in the inference task.

that is a place holder for a conjunctive phrase (e.g., "as a result") explicitly marking the relations involved. After reading the third sentence, a fourth display is presented containing two alternative conjunctions, one representing the relation and the other totally inappropriate in the paragraph context. The subject must press a response key indicating the correct conjunction. Following this response, the incorrect conjunction is erased leaving the correct item on display. Finally, a fifth display is presented, containing a multiple-choice comprehension question. After the subject responds, the correct answer is indicated.

The passages included 4 practice items followed by the test passages. There were three versions of the inference task, each containing a separate set of 20 test items drawn from the original corpus of 64 passages and matched on difficulty level. The difficulty was measured by presenting the first two sentences of the passages to a group of high school subjects along with the conjunctive phrase and asking them to write an appropriate third sentence to complete the passage. The percentage of completions that were semantically appropriate, our measure of passage difficulty, was balanced across test forms. Finally, to obtain a measure of reading speed, the average reading time per word for the third sentence of each passage was obtained.

2.4 Subjects and Experimental Design

In all, ten subjects were trained on one or more of our instructional systems. Five subjects (JS, MG, KG, OD, TS) received training on Speed. Of these, the first four moved on to the Racer System******** and were compared in their acquisition of decoding skills with two subjects (SZ, RF) who did not have prior Speed training. The same four subjects who were trained using Speed and Racer then received training with the Ski-Jump system. A comparison group of three additional subjects (EK, LN, AC) received training on Ski-Jump alone. All of the subjects attended a public high school in Cambridge, Massachusetts. They were selected on the basis of percentile scores on the Nelson-Denny Reading Test Total Score (a composite of vocabulary and comprehension subscores). Their total scores and percentiles were: JS (32, 10%), MG (39, 29%), KG (30, 14%), OD (36, 26%), TS (28, 9%), SZ (18, 4%), RF (28, 17%), EK (28, 14%), LN (40,30%), and AC (41,32%). The ranges of total scores and percentiles were 18-40 and 4%-32%, respectively. The percentile range for subjects in the Speed evaluation study was 9%-29%; for the Racer evaluation, the range was 10%-29% for the group having previous Speed training and 4%-17% for the comparison group of subjects having no previous training; for the Ski-Jump evaluation, the

---

******** The fifth subject TS was unable to continue with the sequence of instructional games.

range was 14%-32% for the comparison group having no previous Speed and Racer training.

Criterion measures were administered before and after each training system. The context priming task was used only in the evaluation of the Ski-Jump system. And, the three subjects tested in the Spring of 1981 (see below) did not receive the word pronunciation and inference tasks at that time.

The bulk of training was conducted over a seven month period, beginning in February of 1982. However, three of the subjects in the Speed evaluation received training with that system in the prior winter and spring. Two of these subjects (JS and MG) participated in the evaluation of the Racer and Ski-Jump systems. They were given two sessions of Speed "refresher" training prior to participation in the Racer evaluation, and were readministered the battery of criterion measures at that time. (The unit detection task was omitted since the subjects showed no drop in Speed performance from the end of their training in the previous spring to the following winter.

3. RESULTS OF TRAINING

3.1 Overview of Methods of Analysis

Results of training were assessed by examining individual performance records for each game, and by analyzing the effects of training on a series of transfer tasks administered before and after each game. These transfer tasks included four tasks specifically related to components undergoing training (the Perceptual Unit Detection task, the Pseudoword and Word Pronunciation tasks, and the Contextual Priming task), and three tasks representing more general reading performance criteria (the Span of Apprehension and Inference tasks, and a measure of Reading Speed derived from the Inference task).

Separate analyses of variance were carried out for evaluating each of the training systems: Speed, Racer, and Ski-Jump. For the Racer and Ski-Jump systems, subjects were divided into two groups, those who had prior training (with the Speed, or Speed and Racer systems) and those who did not. (For the Speed evaluation, no subjects received prior training.) These group analyses of variance, while based upon small samples of subjects, nonetheless allow us to test for the generality of training effects in a subject population. Effects of training are evaluated in these analyses against individual differences in the effects of training -- the training by subject interactions.

The second focus in our analyses is on individual subjects. In this "case" approach, we sought to test the generality of training effects for a given subject over a population of test materials. For this purpose we carried out analyses of variance for individual subjects, using item variances within cells of the design to estimate error variance.

In reporting our results, we will first present records of performance obtained during training for individual subjects. Second, we will present analyses for each of the transfer tasks. For each task, we will first present the group analysis of variance. When there are individual differences in results of training (and the group results are not representative of individuals), the results of analyses for individual subjects will also be presented.

3.2 Speed Training

Complete records of performance during training for each of the five subjects trained using the Speed system are given in Figures 3.1 through 3.5. Each figure presents records of performance for every unit trained, in the order in which the units were introduced. Three indices of performance are plotted for each unit: the number of runs is plotted on the left (where a run stands for an individual Speed game in which as many as 100 target words may be exposed in the window), and the minimum and

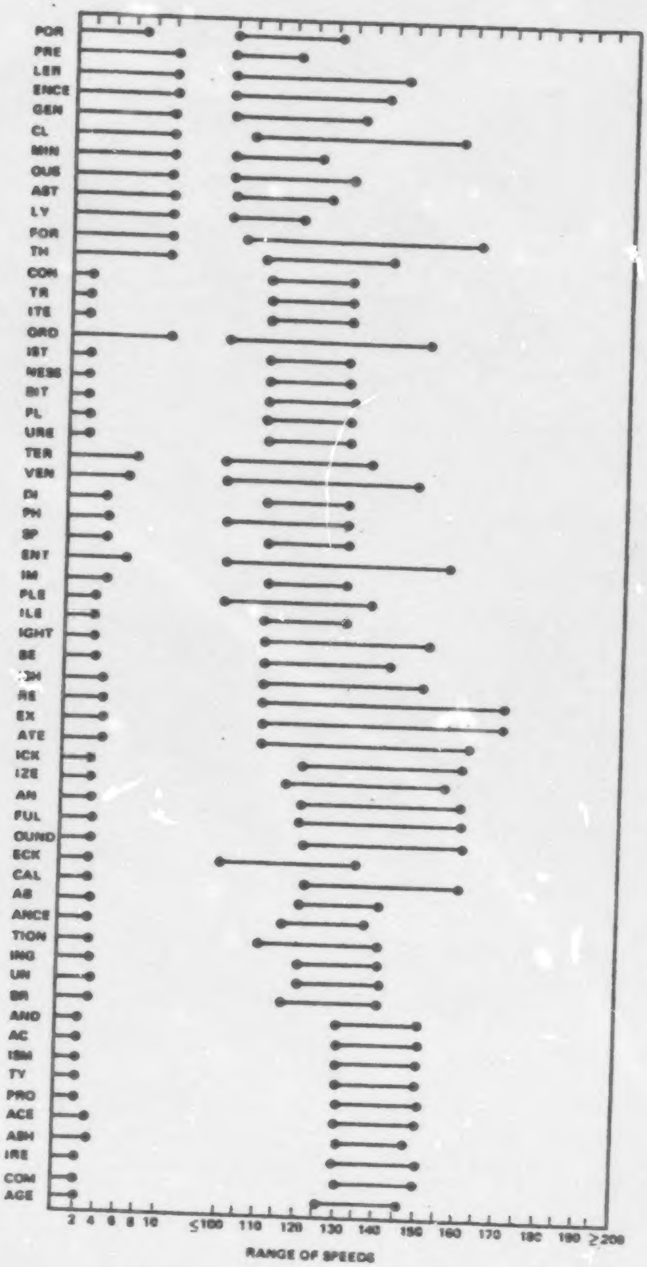


Figure 3.1. Records of Speed training for subject JS. The number of training trials is plotted on the left, and the range of speeds on the right.

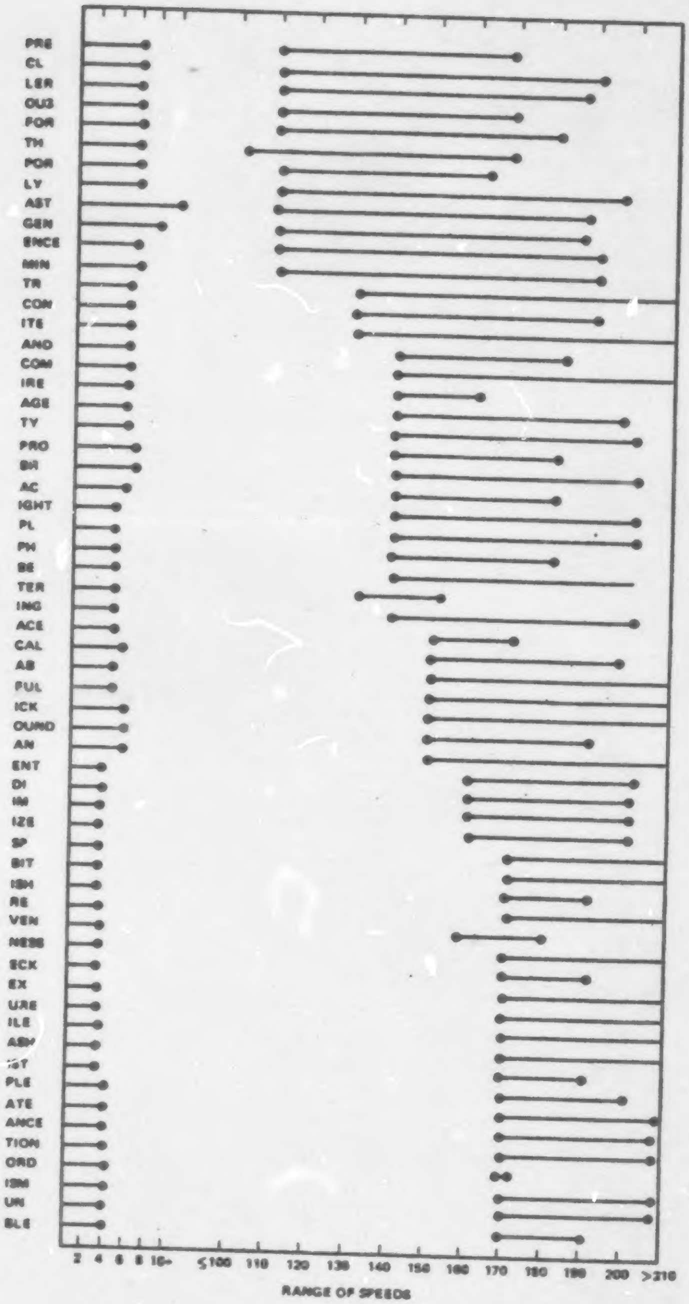


Figure 3.2. Records of Speed training for subject MG. The number of training trials is plotted on the left, and the range of speeds on the right.

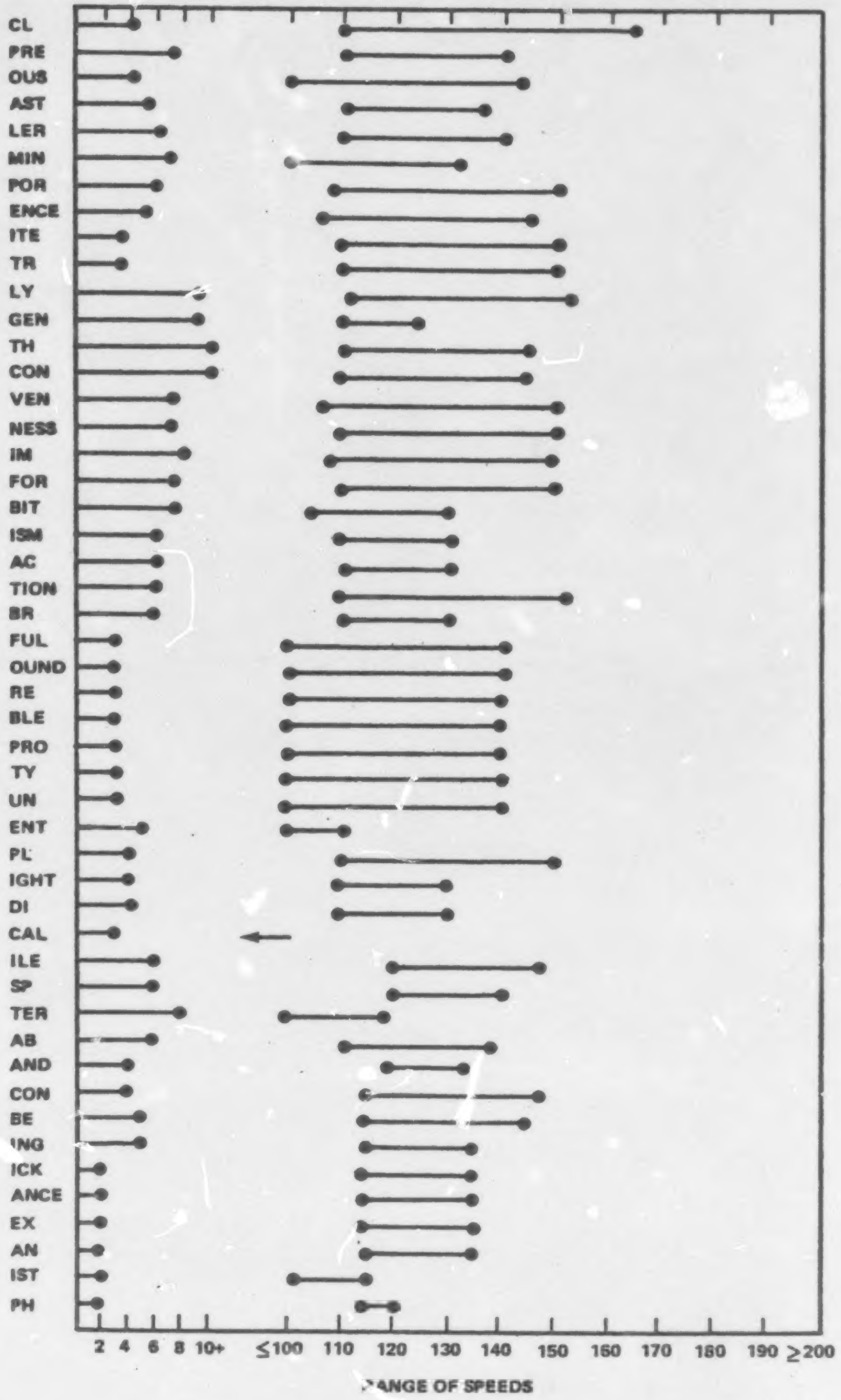


Figure 3.3. Records of Speed training for subject TS. The number of training trials is plotted on the left, and the range of speeds on the right.

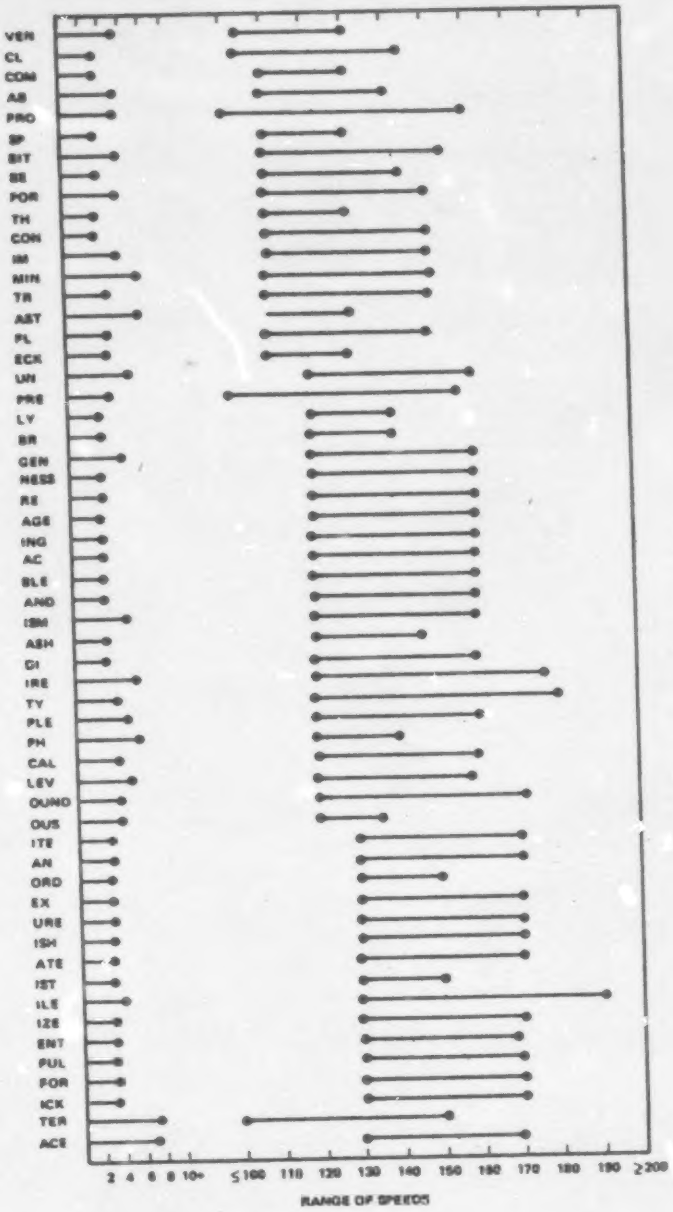


Figure 3.4. Records of Speed training for subject KG. The number of training trials is plotted on the left, and the range of speeds on the right.

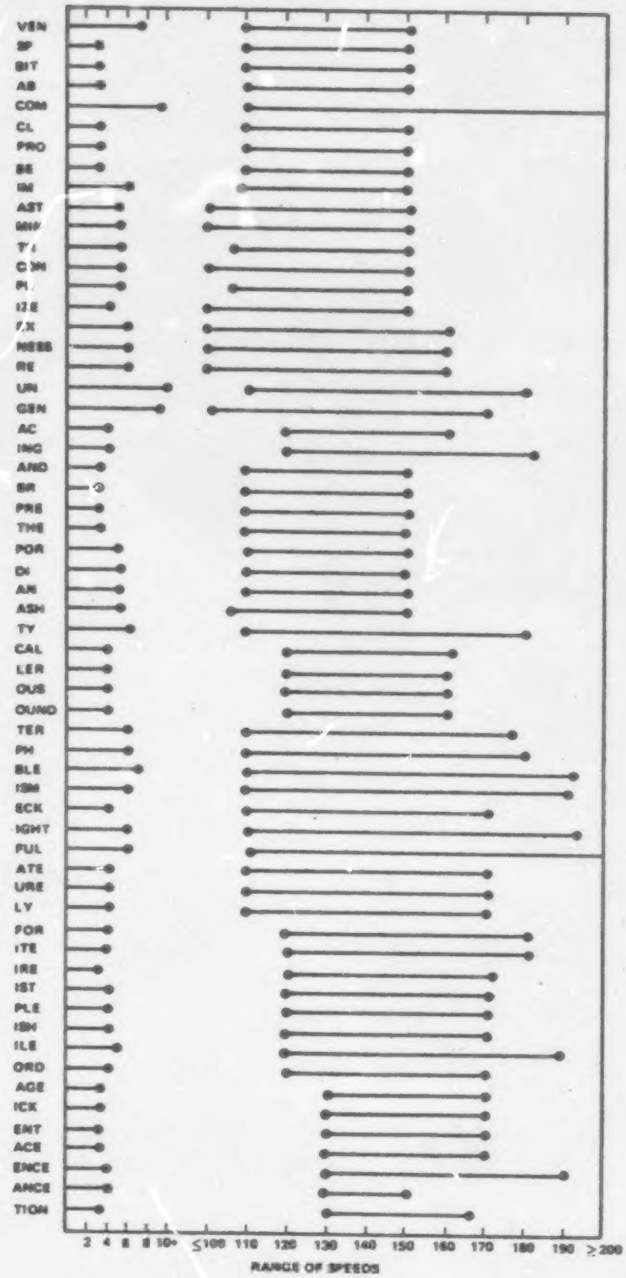


Figure 3.5. Records of Speed training for subject OD. The number of training trials is plotted on the left, and the range of speeds on the right.

maximum speeds attained over those runs are plotted on the right, expressed in words per minute. These three measures should not be viewed as independent. As training progressed, the maximum (criterion) speed required before a unit was considered "mastered" was increased. Thus, mastery of a unit in 2 or 3 runs late in training represents a more demanding accomplishment than it does earlier in training. Our subjects in general required fewer training runs to complete a unit at the end of training than they did initially. By the end of training, subjects JS and TS, for example, required only 2 training runs (the minimum) to master the units. The other subjects ended their training requiring 3-4 runs to complete a unit.

As training progressed, there were pronounced increases in the levels of performance reached on initial runs (the minimum speed) when a unit was newly introduced. For example, subject JS (Figure 3.1) typically had minimum speeds of 100 wpm for units presented early in training (e.g., "por", "pre",..."ly"), but had minimum speeds as high as 130 wpm at the end of training. This transfer of training to as yet untrained units was most dramatic for subjects JS, MG, OD and KG. This strongly suggests that the skill developed in mastering the Speed game is not restricted to those units that have been presented in training, but transfers to untrained units as well. The skill developed thus must involve more general perceptual and attentional skills that can facilitate the rapid detection of previously untrained targets.

Finally, there were large increases in the speeds attained as training progressed. Mastery criteria were set individually and the maximum speeds attained by subjects reflected these individual criteria. Subjects JS and TS reached speeds of 140 wpm, while subjects OD and KG reached speeds of 170 wpm and one subject, MG, reached speeds of 210 wpm and greater. It should be pointed out that a search rate of 200 wpm represents a mean detection latency of 300 msec, a time which includes the motor response time as well as detection time. Such a speed is near the limit of human capability for a task in which individual motor responses are required to every stimulus item in a series.

Transfer of Training

Unit detection task. Improvements in the targeted unit detection skill were evaluated using a task in which subjects' mean RT for detecting multi-letter units was ascertained, under conditions in which the stimulus words (which either contained the units, contained similar units that differed by one letter from the target words, or were dissimilar foils) were followed by a masking stimulus. In the criterion task, unlike the Speed task, the subject completed a response to each word before the next item was presented. Individual detection RTs were obtained for each unit and test word. The initial analysis included training (pretest, posttest) and test word type (targets, similar foils, and dissimilar foils) as factors. Results are presented in Figure 3.6 for mean response latencies. Significant decreases in RT occurred for all stimulus types ($F_{1,4}=19.64$, $p<.01$). There was also a significant interaction of word type and training, with $F_{2,8}=4.37$, $p<.05$. The mean RT for correct target detections decreased from 340 msec before training to 250 msec following training. In the analysis of accuracy of responses, there was a significant effect of word type ($F_{2,8}=44.6$, $p<.001$) and a significant interaction of word type and training ($F_{2,8}=6.04$, $p<.025$). Accuracy was highest for words containing the target unit and for dissimilar foils, and did not change for either of these word types as a result of training. (For targets, it was 92% in both the pretest and posttest, while for foils it was 92% in the pretest and 91% in the posttest.) Accuracy for *similar* foils--words containing a sequence of letters similar to the target unit--was lower than that for the other word types in the pretest (77%), and showed a drop in the posttest (to 63%). This suggests that subjects were lowering their unit detection criterion as a result of Speed training.

Two additional analyses were carried out for trials in which the test words actually *contained* the target units. The first analysis focused on units that appeared in the initial position within test words. Factors that varied in the analysis included (1) effects of training (pretest, posttest), (2) unit set (trained units vs. untrained control units), and (3) unit length (2 or 3 letters). Unit length can also be thought of as a "stand

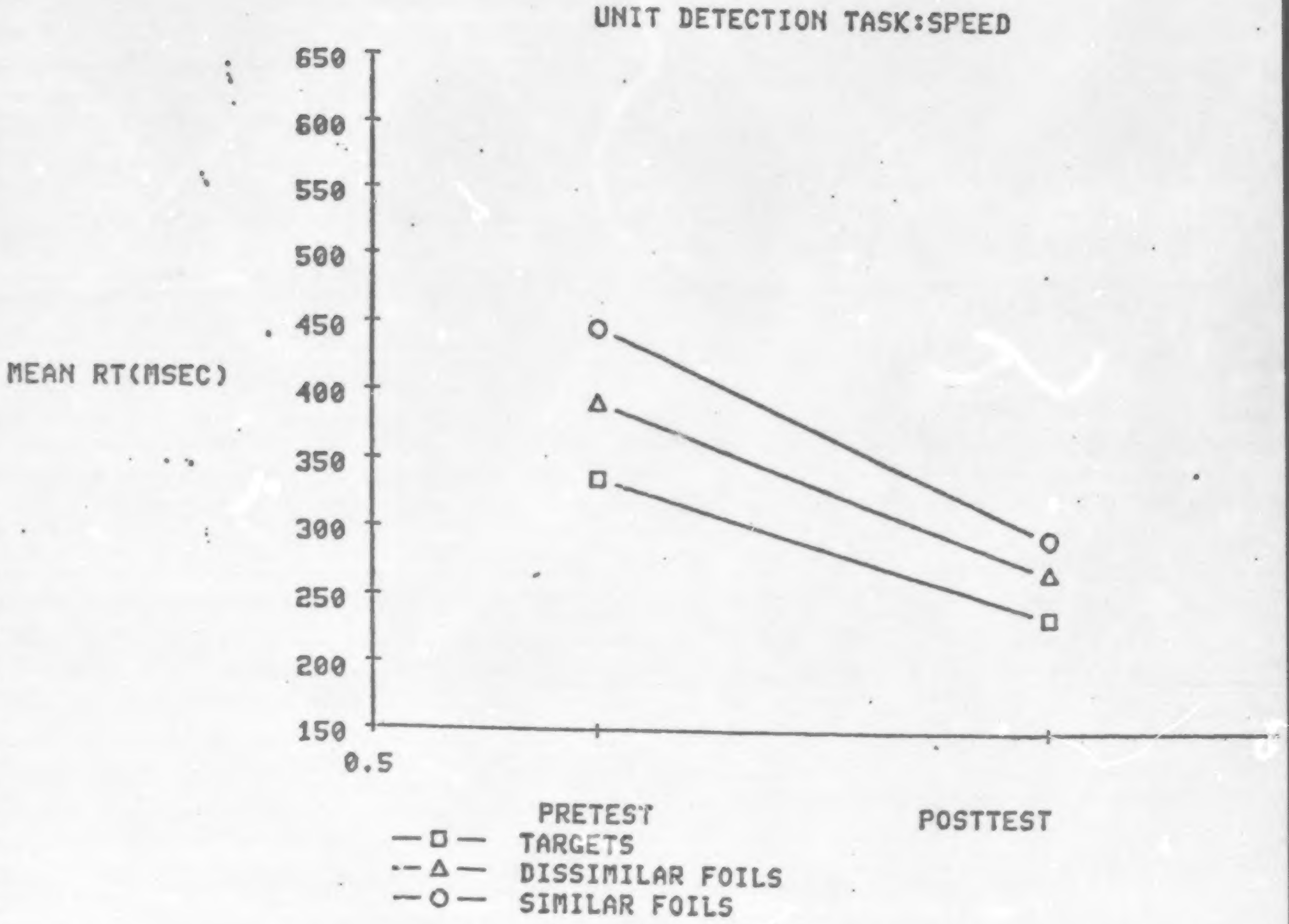


Figure 3.6. Effects of training on Speed on mean RT for detecting multiletter units. Mean RTs are for correct responses.

in" for unit frequency; the mean Mayzner and Tresselt (1965) frequency of 2-letter units was 392, and that of 3-letter units was 57. The mean RT for each of these conditions is shown in Figure 3.7. (The proportion of correct detections varied little from condition to condition, ranging from 87% to 94%.) There was a significant main effect of training on RT for both the trained and untrained units ($F_{1,4}=14.2$, $p <.02$). There was also a significant interaction between unit length and unit set ($F_{1,4}=31.4$, $p <.005$), reflecting a difference in difficulty of 2- and 3- letter units in the trained and untrained sets. These pretest differences primarily reflect sampling variation in unit difficulties for those units assigned to the various conditions. However, whatever the source of these initial differences in unit difficulties, it is clear that following training the unit length and set effects are greatly reduced. (In the analysis, the unit length by unit set by training interaction was significant, with $F_{1,4}=9.1$, $p <.04$.)

The second analysis we carried out focused on the effects of position of a target unit within a test word. In this case, unit length was fixed at two letters. The effects of unit set, unit position, and training on mean detection RTs are shown in Figure 3.8. The mean percents of correct unit detections are plotted in Figure 3.9. There was a significant main effect of training on unit detection RTs ($F_{1,4}=31.4$, $p<.005$), a significant effect of unit position within the test word ($F_{2,8}=17.3$, $p <.001$), and a

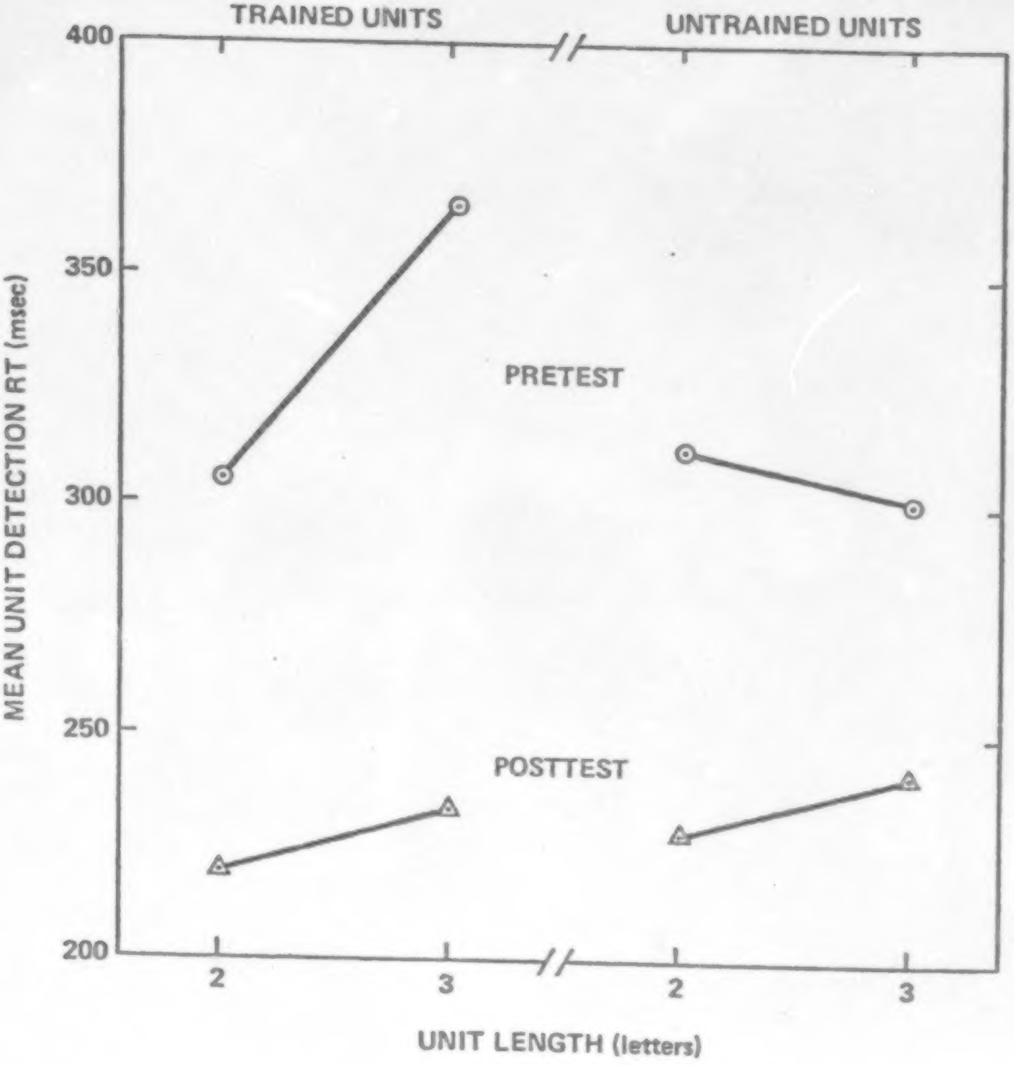


Figure 3.7. Effects of Speed training on mean RT for detecting units that occur at the beginning of target words. Mean RTs are for correct responses.

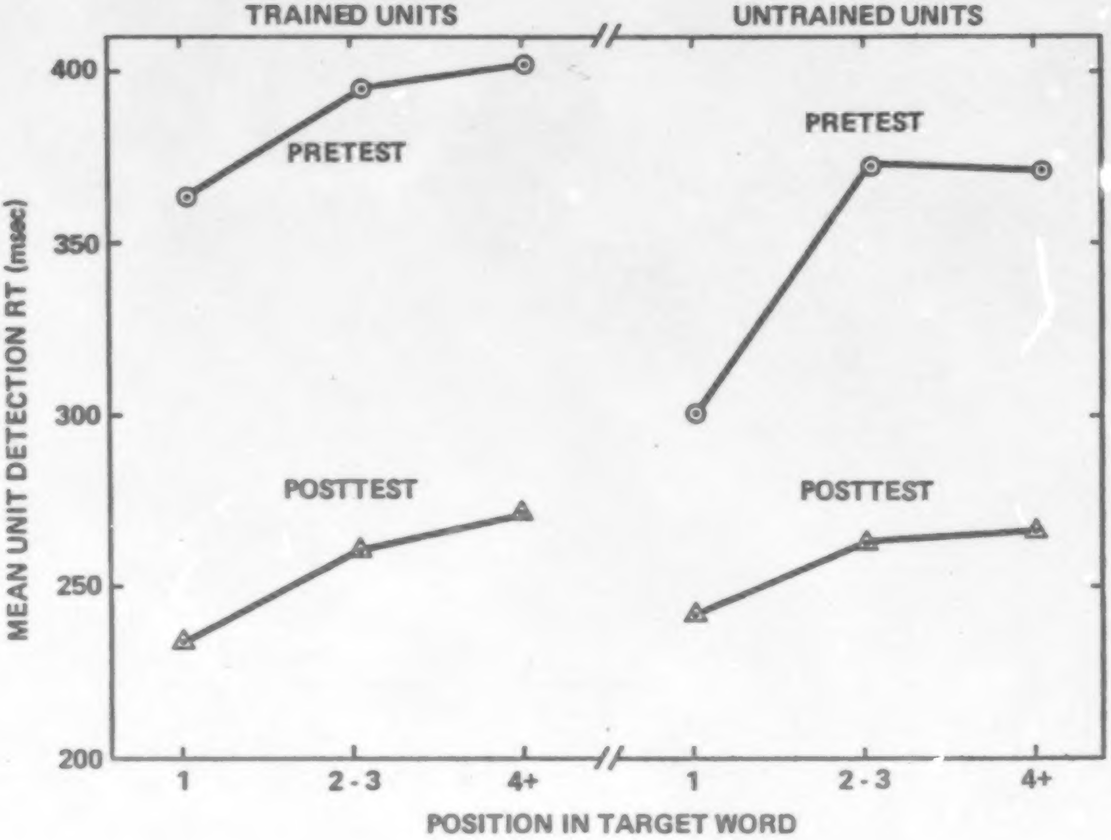


Figure 3.8. Effects of Speed training on mean RT for detecting 2-letter units that occur within target words. Unit positions are the position of the first letter of a unit within the target word. Mean RTs are for correct responses.

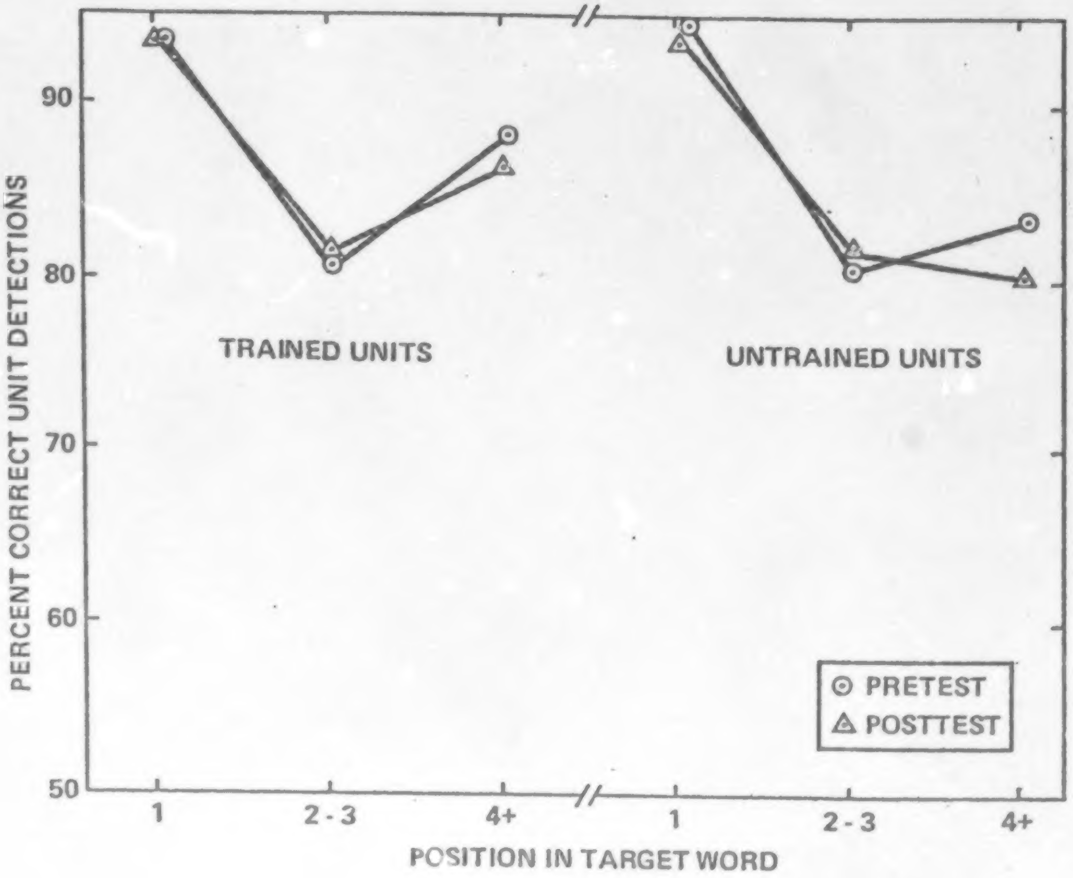


Figure 3.9. Proportions of correct responses in the unit detection task for 2-letter units appearing at varying positions within a target word. Unit positions are the positions of the first letter within the target word.

significant interaction between these factors ($F_{2,8}=6.97$, $p=.018$). Position effects were larger before training than after training. Moreover, the effects of training transferred to the untrained units; there was no significant main effect of unit set (trained/untrained units) in the analysis, nor were there any significant interactions involving this factor. The proportion of correct detections (Figure 3.9) varied as a function of unit position ($F_{2,4}=6.33$, $p=.023$), and there was a significant interaction of position and unit set ($F_{2,8}=19.0$, $p<.001$). (This interaction reflects the difference in performance for units in positions 4 or more letters from the beginning of a test word for units in the trained and untrained sets.) However, there were no significant effects of training in the analysis. Note, finally, that posttraining RTs, which ranged from 233 msec in position 1 to 272 msec in position 4+, are near the limits of human capacity, as they include the motor as well as the detection component. These performance levels were attained for untrained (control) units as well as for units that were used in the training system. Our conclusion is that the skill acquired in the Speed game is a general ability to encode efficiently and accurately the orthographic information present within a word.

Results of the analyses of variance carried out for individual subjects were consistent with those for the group. For every subject there were significant effects of training on RT for detecting units when they occurred within words. And, plots of the results for individuals closely parallel mean RTs shown in Figures 3.7 and 3.8.

Reference group experiment. Four groups of subjects representing varying levels of general reading ability were tested using the unit detection task. The Nelson Denny percentiles for the four groups are: (1) 9-12%, (2) 22-59%, (3) 64-79%, and (4) 94-99%. The task employed was similar to the one used in our evaluation of Speed training except for three features: words were not restricted to 6-9 letters, but covered a wider range from 4 to as many as 12 letters; four letter target units were included; and units were allowed to appear in the final position within target words. Mean reaction times for the four groups have been calculated for each of the conditions represented in Figure 3.7 and 3.8, and allow us to compare performance of our present subjects before and after training with that of the four reference groups. Before Speed training, our subjects had a mean RT of 365 msec for 3-letter and 305 msec. for 2-letter units appearing in the initial positions of words. The corresponding mean RTs for our reference groups for 3-letter units were: (1) 362 msec, (2) 322 msec, (3) 302 msec, and (4) 287 msec, while for 2-letter units they were: (1) 312 msec, (2) 309 msec, (3) 281 msec, and (4) 274 msec. Thus, our subjects pre-test RTs for units in the initial position resembled those of subjects in groups 1 and 2. The same was true for 2-letter units appearing in midword positions. The mean RTs for the four reference groups were: (1) 427 msec, (2) 400 msec, (3) 391 msec, and (4) 383 msec. The mean RT of our trainees in the pretest was 400 msec (for the 2-3 and 4+ positions combined), a value similar to that for the second reference group.

The mean RTs of our trainees following Speed training represent a dramatic improvement over the performance of even the fourth reference group subjects who were at or above the 94th percentile in general reading ability. Our trainees mean posttest RTs for 3-letter and 2-letter units in initial positions were 234 msec and 219 msec, respectively, compared with values of 287 msec and 274 msec for the highest reference group. Likewise, our trainees mean RT for 2-letter units in midword position (2-3 and 4+ position within target words) was 267 msec, which is considerably lower than the 383 msec required by our highest ability reference group. Our conclusion is that training of low ability readers on the perceptual component skill can produce gains in performance that surpass the difference in skill between the highest and lowest reading ability groups tested.

Pseudoword and word pronunciation tasks. There was evidence of transfer of training effects to the performance of the pseudoword and word decoding tasks. Mean pseudoword pronunciation latencies are given in Figure 3.10 and mean percent correct pronunciations are plotted in Figure 3.11. While the mean pronunciation latencies decreased as a result of training,

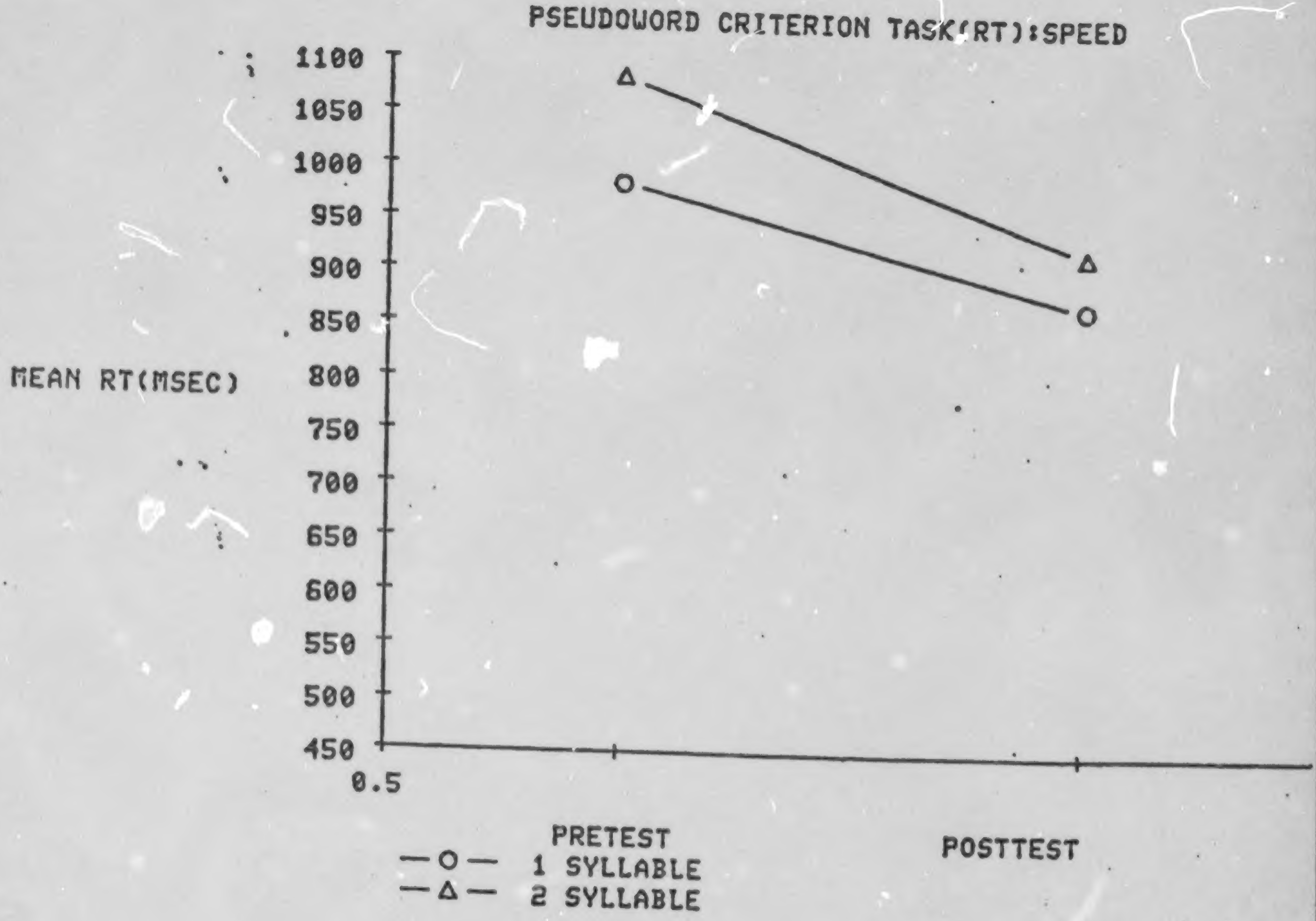


Figure 3.10. Effects of training on Speed on mean RT for pronouncing pseudowords. Mean RTs are for correct responses.

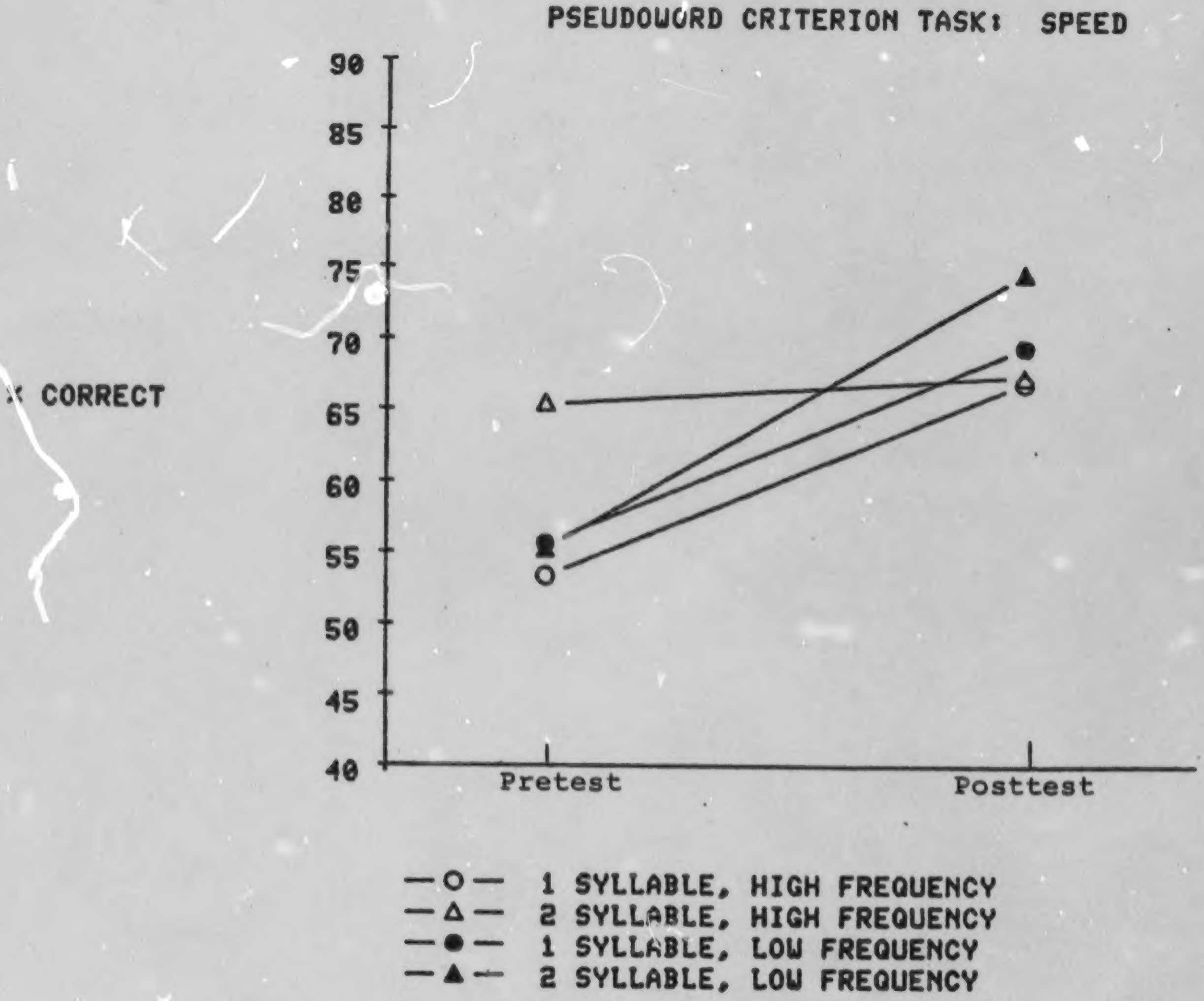


Figure 3.11. Effects of training on Speed on the proportion of current pronunciations in the pseudoword pronunciation task.

101

the effect was not statistically reliable. There were, however, significant effects of training on subjects' accuracy of pronunciation. Increases in accuracy amounting to 14% and 20% were found for one and two syllable words derived from low frequency words. For pseudowords derived from high frequency words, there was an increase in accuracy of 14% for one syllable items, but only a 2% increase for two syllable items, which were initially easier for our subjects to pronounce. The analysis of variance resulted in a significant frequency by training interaction ($F_{1,4}=15.1$, $p=.02$), a marginally significant main effect of training ($F_{1,4}=5.76$, $p=.07$), and a marginal triple interaction of training, frequency, and syllable length ($F_{1,4}=6.70$, $p=.06$).

There were individual differences in the effects of training on pseudoword pronunciation. Two of five subjects (OD and KG) showed significant *reductions* in RT ($F_{1,138}=47.4$, $p<.001$ and $F_{1,139}=56.7$, $p<.001$, respectively), accompanied by increases in or maintenance of high accuracy (80% to 95% for the first subject, and 89% to 91% for the second). Two other subjects (JS and MG) showed *increases* in latency ($F_{1,296}=4.55$, $p=.03$ and $F_{1,296}=17.4$, $p<.001$), accompanied by *increases* in accuracy. For example, initially the first of these subjects, JS, could correctly pronounce only 13% of our list of 4-6 letter pseudowords; he was essentially guessing on the basis of one or two initial letters that he could perceive within the brief (200 msec) stimulus exposure employed in the task. Following Speed training, his accuracy improved to 41%. While still having great difficulty with decoding (his RT was as long as 1 1/2 seconds), he was now able to base his efforts on accurately encoded orthographic information. The second subject, MG, initially pronounced 62% of the pseudowords correctly with an average latency of 807 msec. Following training, he was correct on 78% of the items, with an average latency of 1022 msec. The final subject, TS, showed no change in latency or accuracy on the pseudoword pronunciation task.

The two Speed subjects (OD and KG) who were trained in the winter of 1982 were also administered a word pronunciation task before and after their training. Both of these subjects showed significant reductions in vocalization onset latency, with $F_{1,144}=29.3$, $p<.001$ for subject OD and $F_{1,139}=112.8$, $p<.001$ for subject KG. Both subjects showed increases in accuracy as well, the first from 85% to 96% and the second from 90% to 98%. Thus, training in the perceptual skill has an influence on performance of the higher level word/pseudoword decoding task. This "forward transfer" of skill is presumably due to the effect of training on the availability of orthographic units upon which decoding is based.

Span of apprehension task. There was evidence of transfer of acquired skill in perceptual encoding to a measure of span of apprehension, but only for some subjects. Increases in mean visual span for the five trainees are shown in Figure 3.12. The effects of training were marginally significant, with $F_{1,4}=4.99$, $p=.09$. Increases in visual span were accompanied by a reduction in mean response latency (shown in Figure 3.13), but again these effects of training for the group as a whole were not significant ($F_{1,4}=3.47$, $p=.14$). Out of the five subjects, four showed significant improvement in either response latency or visual span. Subject KG showed an increase in span in the no context condition from 12.9 to 15.7 letter spaces ($t_{78}=2.62$, $p<.01$). She showed no change in performance in the context condition, but her span of 14.2 letter spaces prior to training (and 14.4 following training) was near asymptote for that task. She showed no significant changes in response latency as a result of training. Subject JS showed very small measured spans in both the context (5.6) and no context (7.9) conditions prior to training, but showed a significant increase to 9.7 letter spaces for the context condition (for the no context condition, his span remained the same at 8.1 letter spaces). Both the main effect of training and the context by training interaction were significant, with $F_{1,152}=9.75$, $p<.01$ and $F_{1,152}=8.02$, $p<.01$, respectively. JS also showed a significant reduction in response latency, from an average of 2530 msec prior to training to 1446

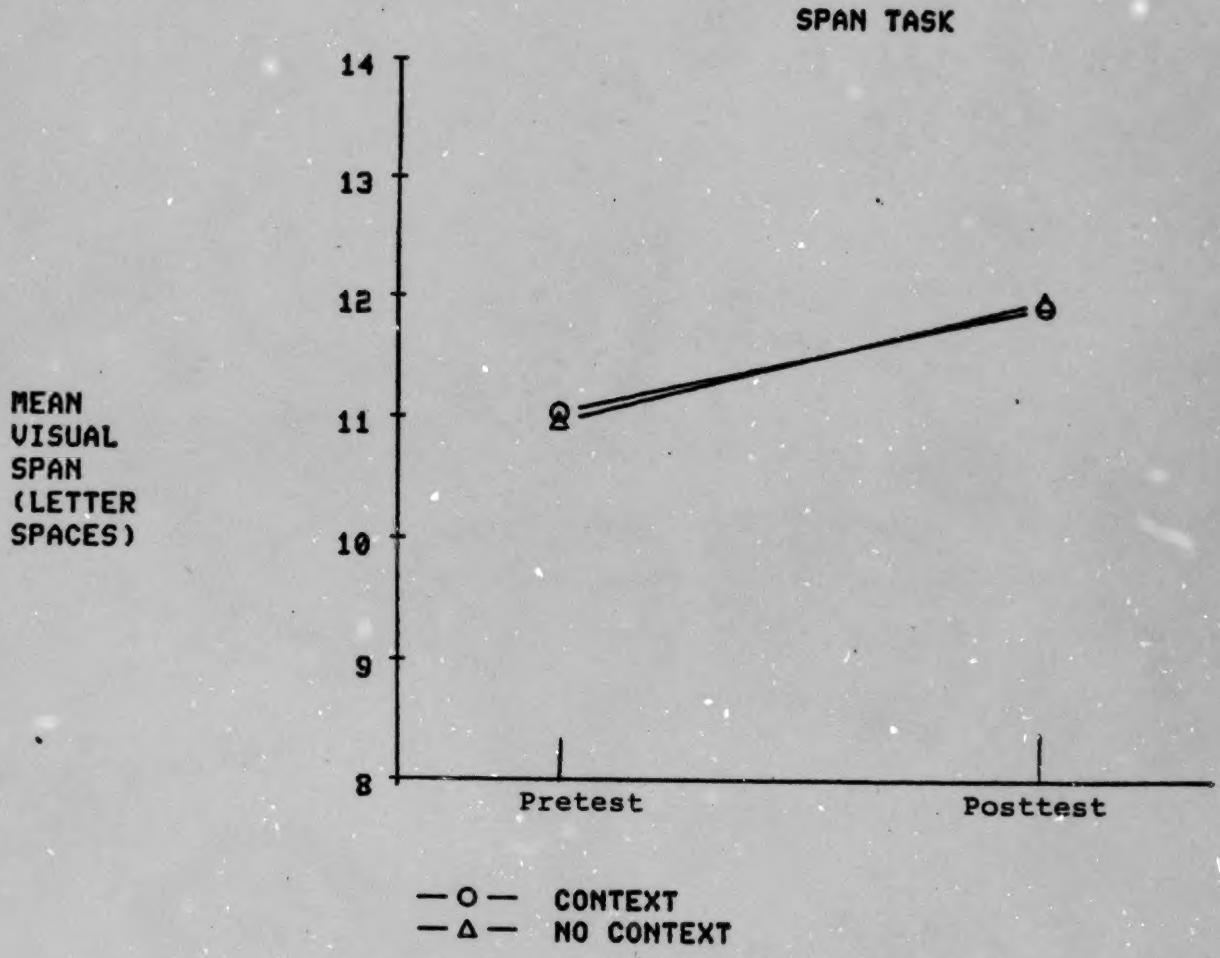


Figure 3.12. Effects of Speed training on the span of apprehension.

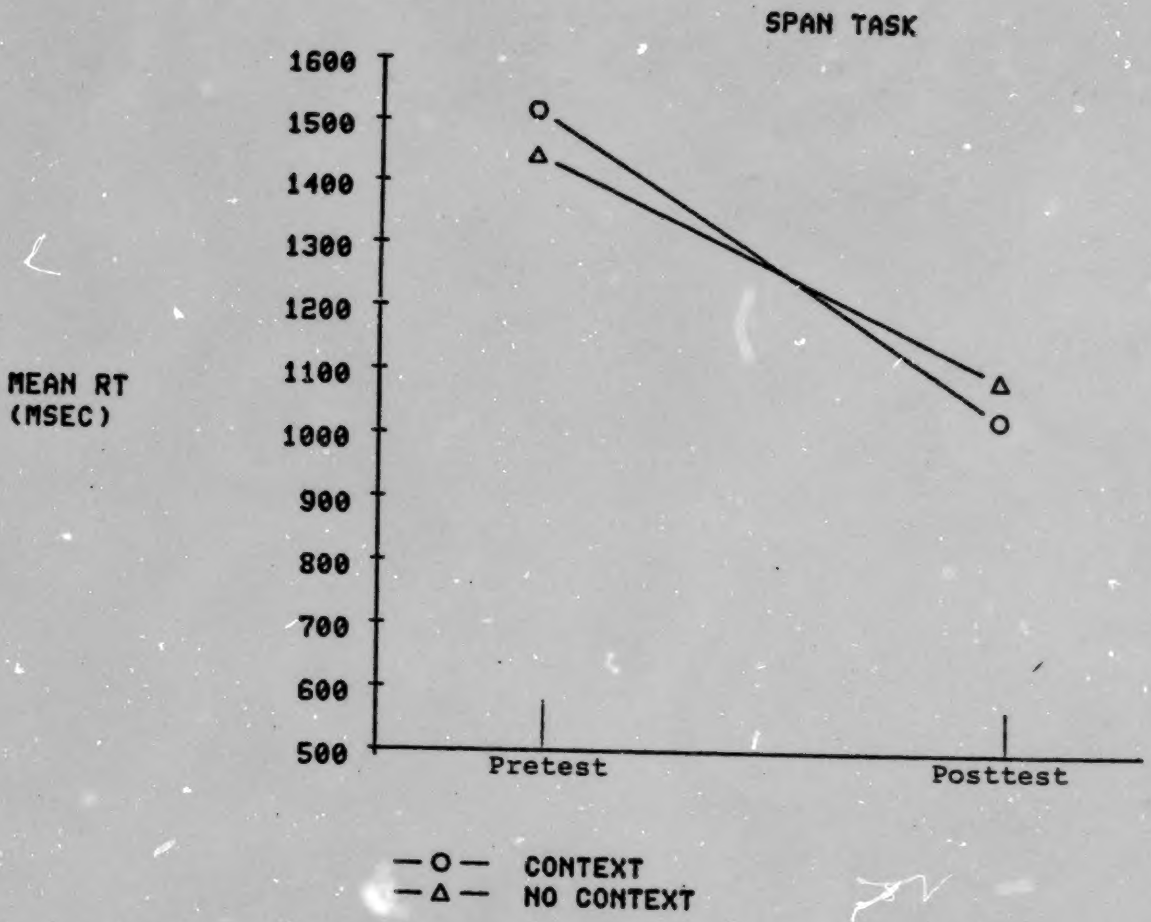


Figure 3.13. Effects of Speed training on mean RT in the span of apprehension task.

msec following training, with $F_{1,136}=11.22$, $p<.01$. Two of the remaining three subjects showed significant reductions in response latency. Subject TS showed no change in visual span (an average of 10.6 prior to and 10.9 letter spaces following training), but showed a significant reduction in latency from 1824 msec to 1006 msec, with $F_{1,148}=17.6$, $p<.01$. Subject MG also showed no significant change in span (14.9 prior to and 15.9 letter spaces following training), but showed a significant reduction in latency, from 964 msec to 772 msec, with $F_{1,156}=4.05$, $p<.05$. His pretest span was near the ceiling for this task, however, so there was little room for him to improve.

Inference task. Only the two subjects trained in the winter of 1982 received pre- and posttests on the inference task. Of these subjects, one (KG) showed a significant improvement in accuracy in judging appropriateness of connectives, with 40% correct prior to training and 65% following training ($t=1.89$, $p<.05$). Neither subject showed any significant change in reading rate.

Summary. Training on the Speed system resulted in significant improvement for all subjects on criterion tasks designed to assess the effects of instruction on the perceptual skills targeted for development. The improvements in unit detection performance extended to test units that were matched with those used in training but were not actually used in the

Speed system. The perceptual skill developed appears therefore to be more general than a set of specific unit detectors, and involves a general improvement in the ability to encode orthographic information within a word. In addition, there is evidence for a causal link between a perceptual skill and skill in decoding (cf. Frederiksen, 1981). The development of the perceptual skill had a significant impact on accuracy of decoding, although this improvement was accompanied by an increase in RT for some subjects. The Speed system thus serves to develop an ability to encode effectively the orthographic information necessary for accurate decoding, an important prerequisite for later training aimed at automating word decoding. Evidence for transfer of Speed training effects to a more general criterion measure, the span of apprehension, supported a tentative conclusion that the acquired perceptual encoding ability can have an impact on the amount of information encoded within a fixation or on the time needed to process such information, even without subsequent training to allow practice in the application of the perceptual skill in the performance of actual reading tasks.

3.3 Racer Training

Complete records of training on the Racer system are presented for each subject in Figures 3.14 through 3.19. Of these subjects, the first four received prior training using the Speed game, and the last two (SZ and RF) had no prior training.

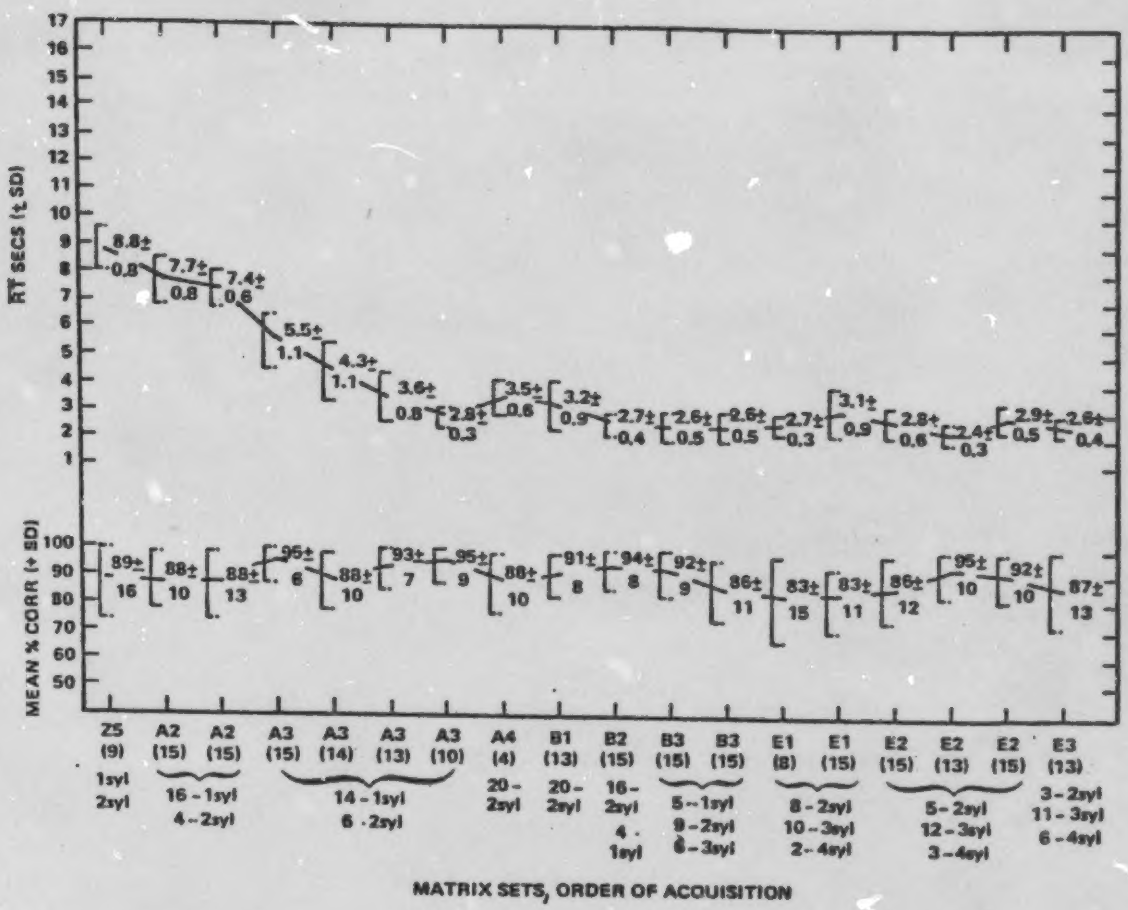


Figure 3.14. Records of training using Racer for subject KG. The top curve gives the mean RT and SD for completing a matrix, and the bottom curve gives the mean percent of correct responses in the Sound Trap portion of the game.

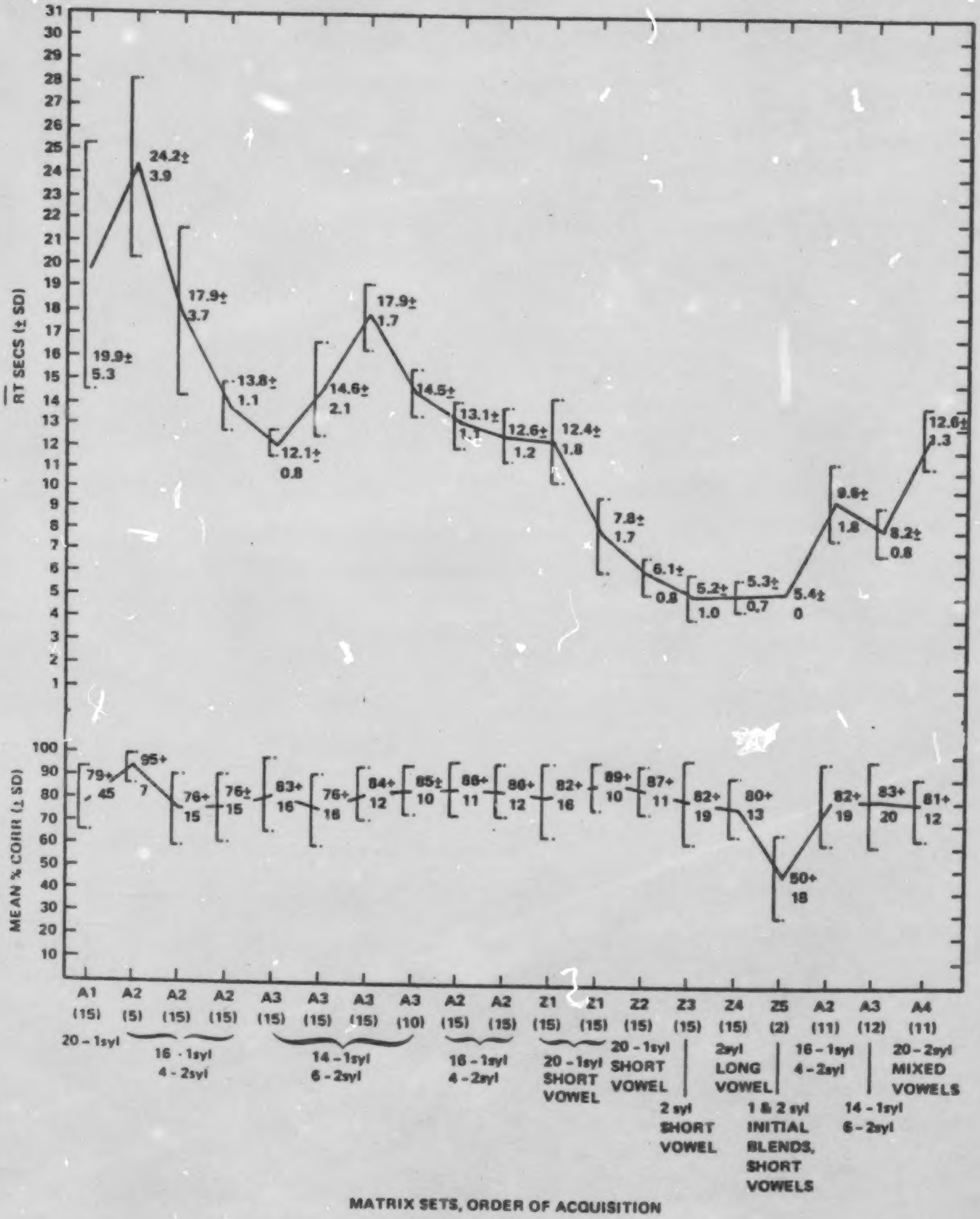


Figure 3.15. Records of training using Racer for subject JS. The top curve gives the mean RT and SD for completing a matrix, and the bottom curve gives the mean percent of correct responses in the Sound Trap portion of the game.

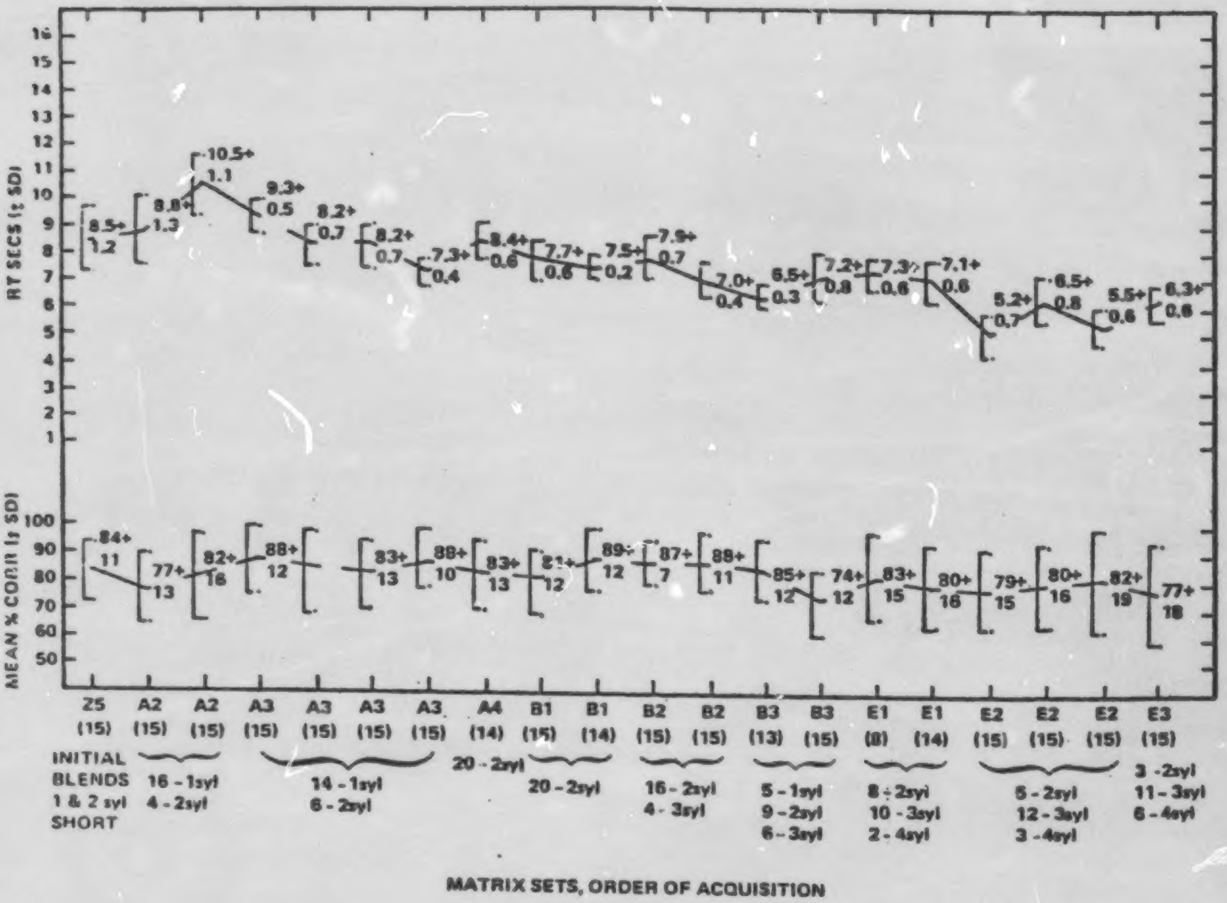


Figure 3.16. Records of training using Racer for subject OD. The top curve gives the mean RT and SD for completing a matrix, and the bottom curve gives the mean percent of correct responses in the Sound Trap portion of the game.

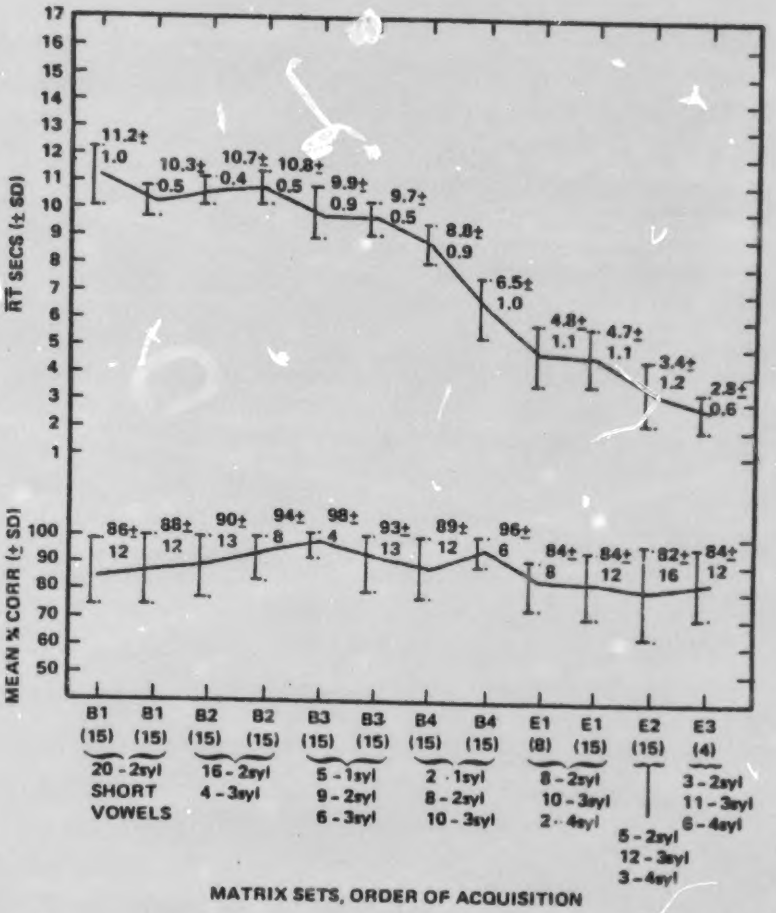


Figure 3.17. Records of training using Racer for subject MG. The top curve gives the mean RT and SD for completing a matrix, and the bottom curve gives the mean percent of correct responses in the Sound Trap portion of the game.

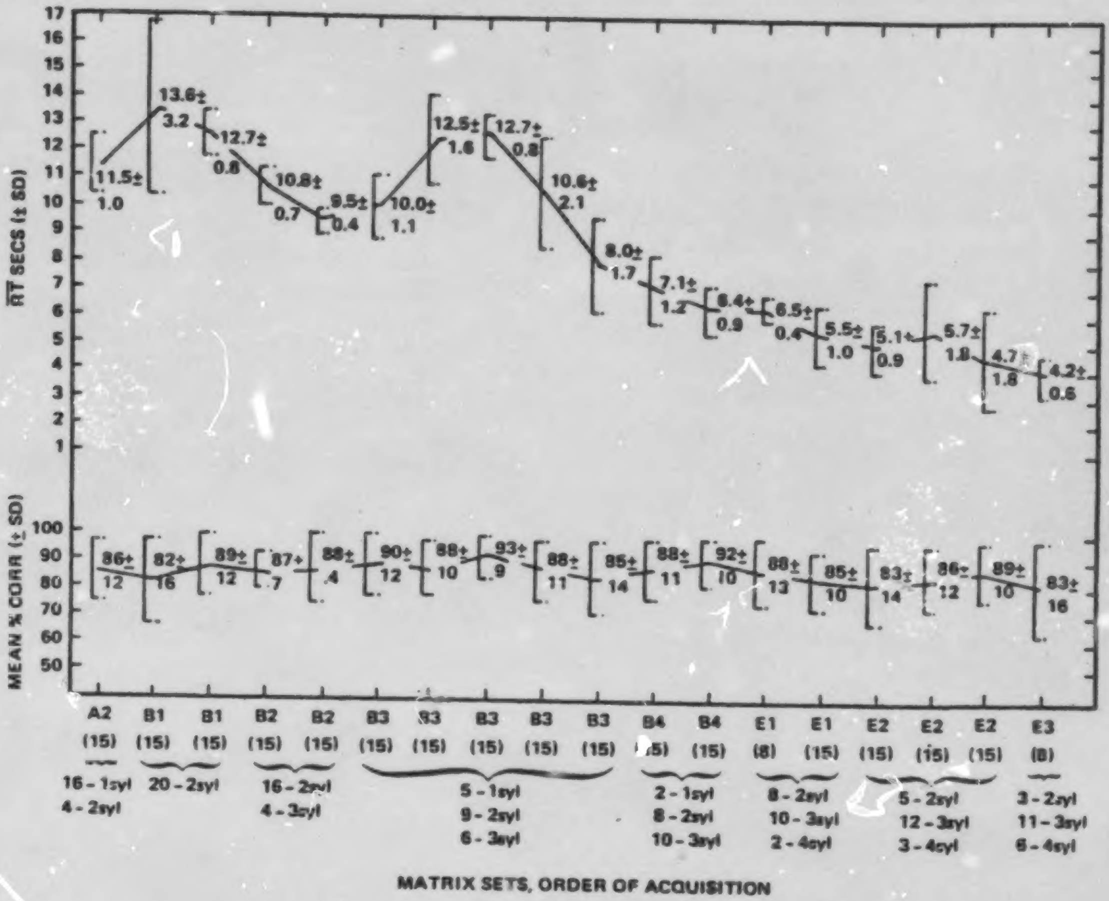


Figure 3.18. Records of training using Racer for subject SZ. The top curve gives the mean RT and SD for completing a matrix, and the bottom curve gives the mean percent of correct responses in the Sound Trap portion of the game.

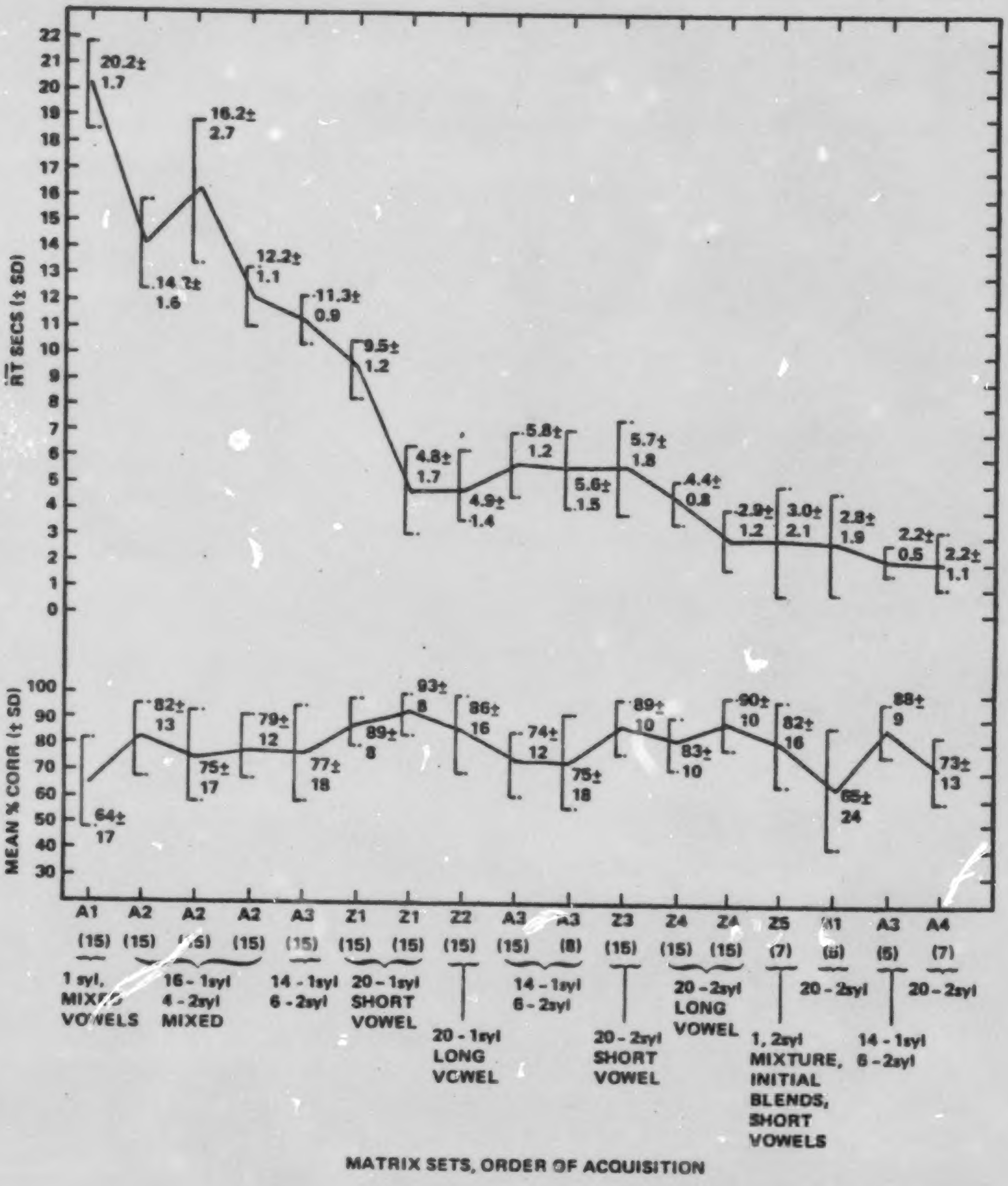


Figure 3.19. Records of training using Racer for subject RF. The top curve gives the mean RT and SD for completing a matrix, and the bottom curve gives the mean percent of correct responses in the Sound Trap portion of the game.

Two dependents are plotted for each matrix set in the training sequence: the mean RT[*********] for completing matrices of a single type within a given session, and the mean percent correct in Sound Trap, averaged for the same sets of matrices. Standard deviations for both measures are also given. The matrix sets are listed on the abscissa, in the order in which they were presented. For each matrix set, the number of training trials is indicated in parentheses, along with their composition.

Some general features of these training records are worth emphasizing. For all subjects over the course of training, there was a gradual reduction in mean RT for reading the words in a matrix, despite the increase in difficulty of words within successive matrices. Some subjects appeared to reach their asymptotic level of performance earlier than did others. Subjects KG and RF (Figures 3.14 and 3.19) appear to have reached their fastest speeds after 7 matrix sets were completed. More typical were subjects OD and MG (Figures 3.16 and 3.17), who were still showing improvement at the end of the training sequence.

The performance gains of our subjects were all substantial, although there are individual differences in decoding ability that are reflected in their records of performance. Four of the

---

[*********] The RT for a matrix is the sum of vocalization onset latencies for the twenty words in a matrix.

subjects (Figures 3.14, 3.16, 3.17 and 3.18) had no difficulty in progressing through the standard series of matrices. The first three of these subjects showed no substantial reversals in performance, and ended training with mean RTs for matrices composed largely of 3- and 4-syllable words that were substantially lower than their initial RTs for matrices of primarily one-syllable words. Subject KG, for example, required an average of 8.8 sec on the first matrix set, and only 2.6 sec on the final set. Subject SZ (Figure 3.18) showed some reversals in performance, but also successfully mastered the standard series of matrices. His final mean RT was 4.2 sec, compared with an initial RT of 11.5 sec. Two of the subjects, JS (Figure 3.15) and RF (Figure 3.19), had more difficulty with the Racer task. Each began training with the standard series of matrices, but was moved into a special series of matrices (the Z-series) when improvement in performance on the standard (A) series was too slow. Subject JS, for example, after having completed a total of 140 trials with A-series matrices, was still taking nearly 15 seconds on the average to complete a matrix. The Z-series matrices were initially designed to help him past this hurdle. These matrices were built around simple, consistent phonic principles (e.g., all long vowels, or all short vowels, or consonant blends). JS showed rapid progress when he began working with the consistent matrices of the Z-series, and his performance on the A-series matrices showed marked improvement when that series was reintroduced. This strategy of building skill through the introduction of consistent matrices was applied to another subject, RF (Figure 3.19), as well. Subject RF also showed immediate gains in performance when the Z-series matrices were introduced, dropping from a mean of 9.5 sec for matrix set A3 to 4.8 sec on the initial Z-series matrices. He reached a level of 3 sec at the end of the Z-series, and continued to perform at that level after advancing to the Standard B-series.

Finally, a comment must be made concerning the levels of performance reached at the end of training. Mean RTs in the neighborhood of 3-4 seconds for a matrix of 20 words imply mean onset latencies for individual words of 150-200 msec. Our subjects were not recognizing words in this absurdly short interval; rather they were using a strategy they developed for keeping ahead of the computer "horse", while still pronouncing the words fully and correctly. Some subjects appeared to begin with a pronunciation of the "initial" consonant or vowel, and draw out that pronunciation while decoding the rest of the word. Another technique was to preface their vocalization with some extraneous sound ("ah") that would trip the voice interface and stop the computer horse. These strategies could only work to a point, for if there was any break in the vocalization, the computer would proceed with the next item, and the subject would quickly fall behind. Subjects thus learned that smooth and correct vocalizations were required if they were to pass Sound

Trap. Our subjects were generally able to maintain high accuracy in Sound Trap (e.g., 80% - 90% correct), even at their final rates of speed.

Transfer of Training

Pseudoword and word pronunciation tasks. The effectiveness of Racer in training efficient decoding was evaluated by testing subjects' speed and accuracy in pronouncing test lists of pseudowords and words, before and after training. Mean pseudoword vocalization latencies obtained before and after training are shown in Figure 3.20 for the groups of subjects who did or did not receive prior Speed training. The accuracy of pronunciation is shown in Figure 3.21. There were significant reductions in RT following training, with $F_{1,4}=8.99$, $p<.04$ for the main effect of training. There was also a significant interaction between syllable length and training, with $F_{1,4}=10.7$, $p<.03$. Prior to training, two-syllable words took, on the average, 245 msec longer to decode than one-syllable words. Following training, there were no differences in latency associated with variations in syllable length. Subjects in the group given prior Speed training had shorter pronunciation latencies in the pretest (798 msec) than did those with no prior training (987 msec). The two subjects in the latter group also showed larger effects of Racer training in the pseudoword pronunciation task. However, the number of subjects in the group

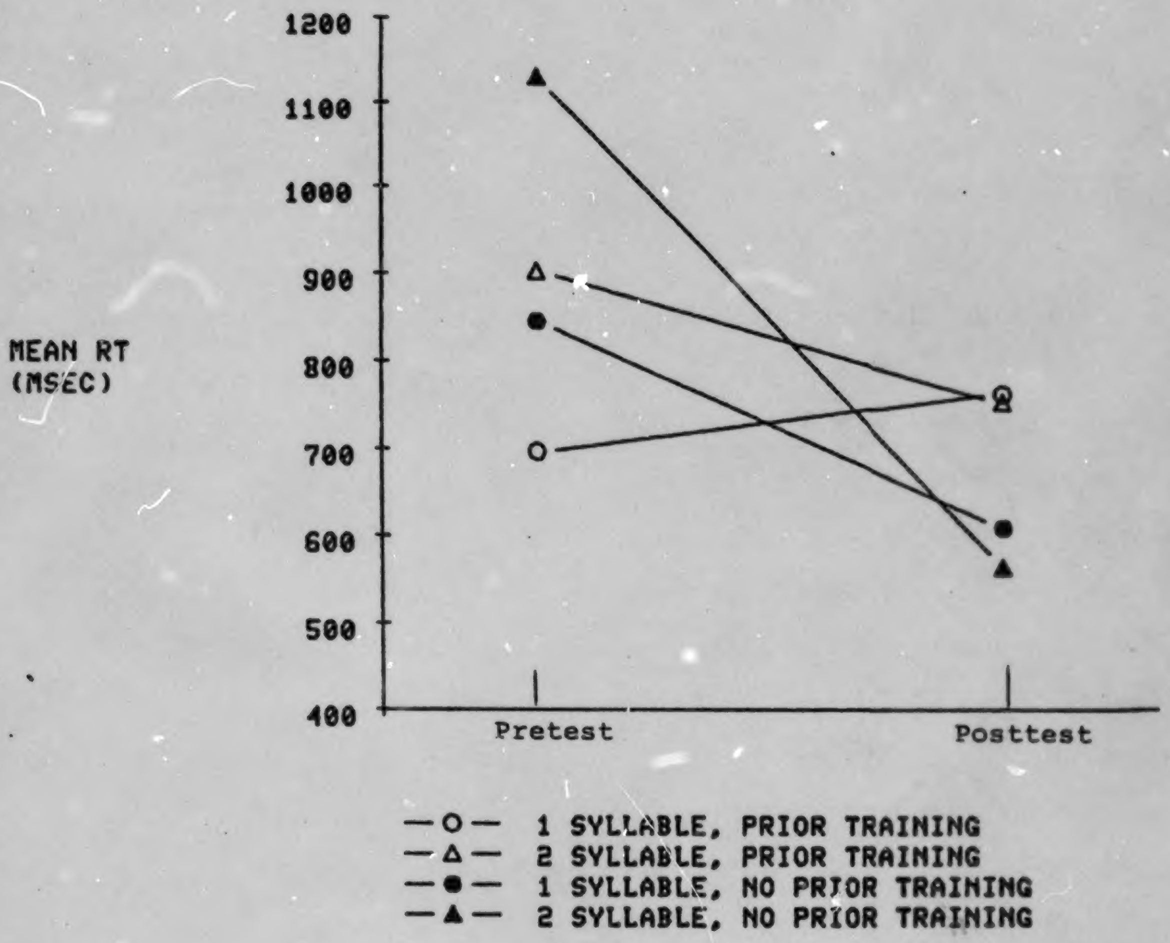


Figure 3.20. Effects of Racer training on mean pseudoword vocalization onset latencies for correct pronunciations. Subjects in the prior training group completed training on Speed before beginning Racer training.

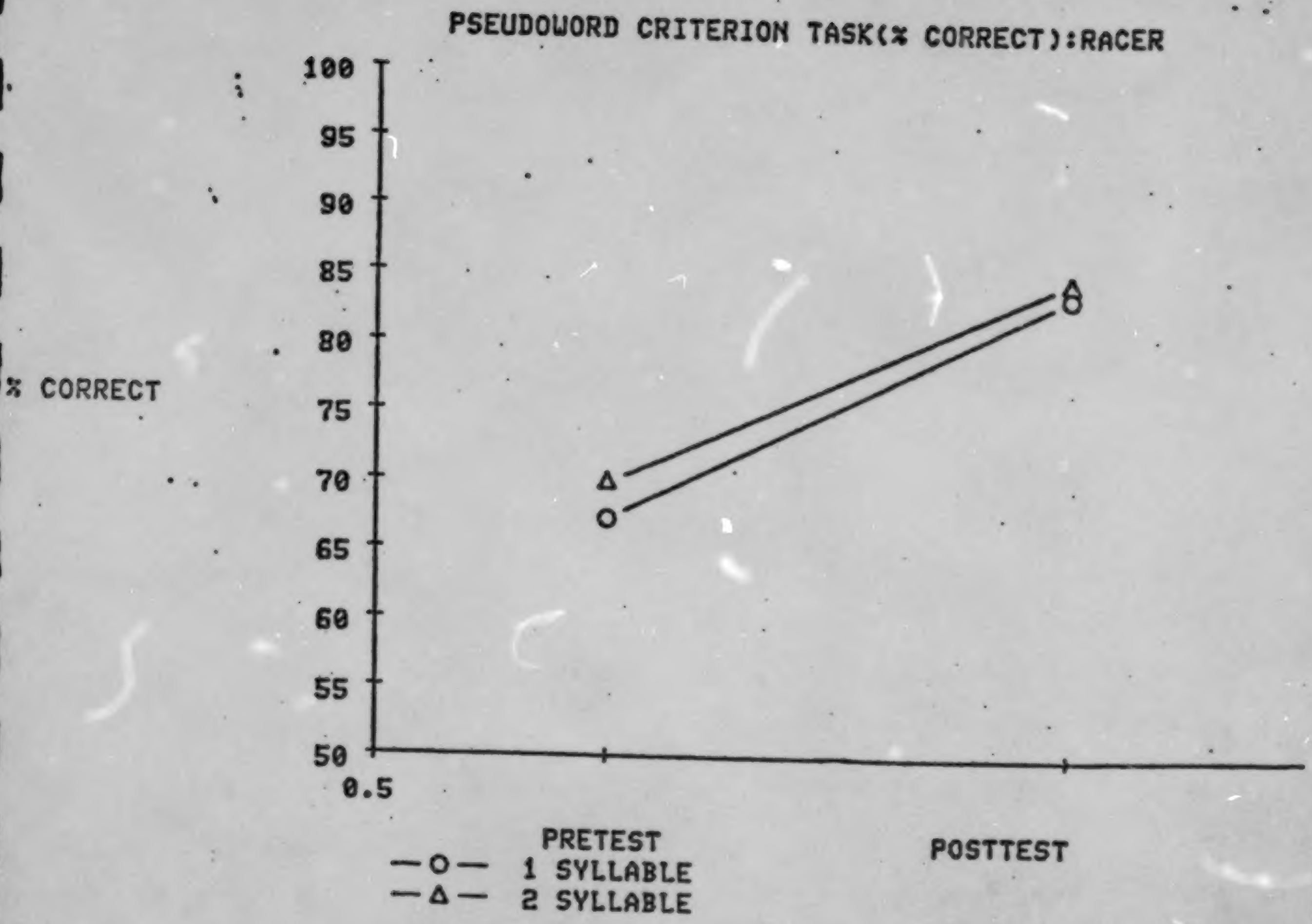


Figure 3.21. Accuracy of pseudoword pronunciation before and after training on Racer. Subjects in the prior training group completed training on Speed before beginning Racer training.

was small and the group by training interaction was only marginally significant ($F_{1,4}$ =5.60, p=.077).

The decreases in pronunciation RT as a result of training were not the result of subjects' acceptance of more pronunciation errors. In Figure 3.21 the mean percent of correct pronunciations for the two groups combined is shown. (There were no group differences in pronunciation accuracy.) Training actually led to a 15% increase in accuracy; however, these effects of training on accuracy were not statistically significant. In the analysis of each subject's responses, four of the six showed substantial improvement in accuracy, and two subjects, whose accuracies were above 90%, showed no further improvement. The final accuracy levels ranged from 69% for subject JS to 90% for subject OD.

Results for the word pronunciation task were similar, and are shown in Figures 3.22 and 3.23. Following training, an average latency of 515 msec was obtained, and the difference in latencies for one- and two-syllable words was reduced from 194 msec before training to 66 msec following training. The main effect of training on mean RT was significant, with $F_{1,4}$ =10.87, p=.03. Again, subjects given prior Speed training had shorter RTs in the pretest than did subjects with no prior training, and their RT was less influenced by the syllable length of the target word. Following training, there were no differences in

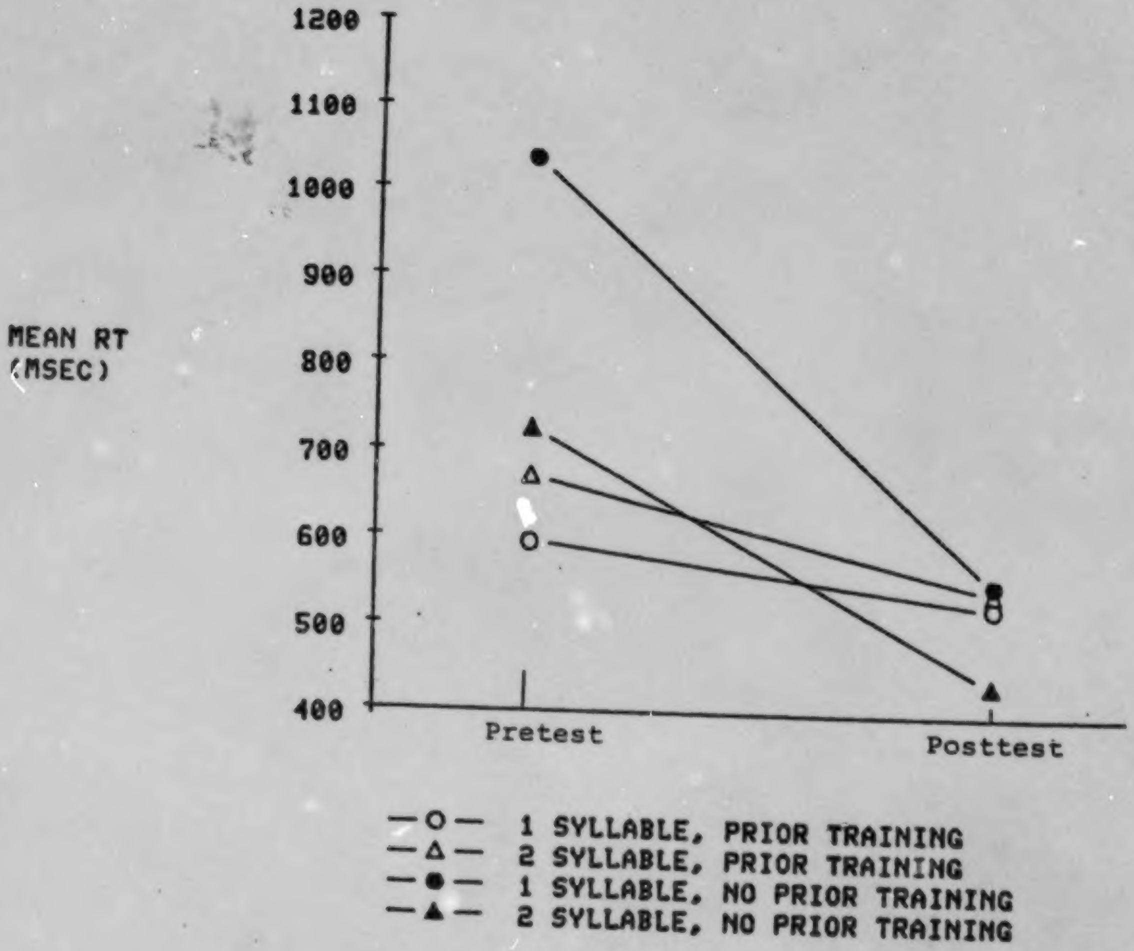


Figure 3.22. Effects of Racer training on mean word pronunciation latencies for correct responses. Subjects in the prior training group completed training on Speed before beginning Racer training.

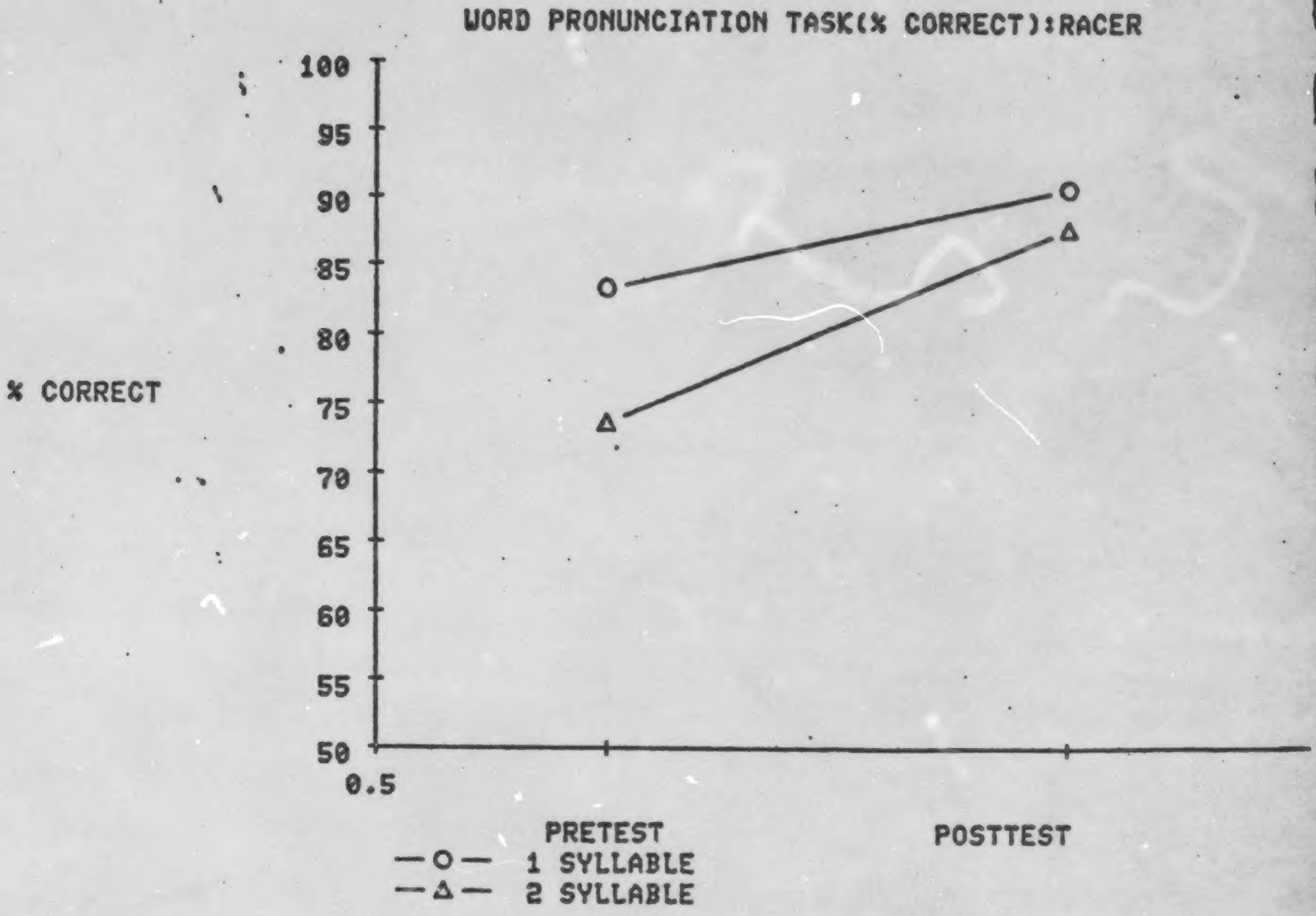


Figure 3.23. Accuracy of word pronunciation before and after training on Racer. Subjects in the prior training group completed training on Speed before beginning Racer training.

performance for these two groups. However, neither the group by training ($F_{1,4}=3.93$, $p=.12$) nor the group by syllable length ($F_{1,4}=4.14$, $p=.11$) interaction was significant. There were also increases in accuracy which are shown in Figure 3.23. The effects of training on accuracy were not significant, however ($F_{1,4}=3.89$, $p=.12$). In the analysis for individual subjects, five of the six subjects showed significant reductions in mean RT for pronouncing words, and all subjects who were not at asymptote prior to training showed increases in accuracy as a result of training. The pre- and post-training accuracy percentages for each subject were: 42-72 (JS), 78-89 (MG), 98-98 (KG), 96-94 (OD), 80-88 (SZ) and 76-94 (RF). Our conclusion is that Racer training not only enabled our subjects to develop highly automatic procedures for decoding words, it enabled them to develop greater accuracy in their decoding as well.

Unit detection task. We also evaluated the effects of Racer training on development of the perceptual skill addressed in the Speed game. We found that Racer training, which did not provide any explicit feedback concerning the perceptual skill, was not effective in developing this subskill, even though it is implicitly involved in word decoding. The pertinent results are shown in Figure 3.24. Mean unit detection latencies did not change as a result of Racer training. However, there were significant effects attributable to item type (target, similar and dissimilar foils) and to group (prior Speed training or no

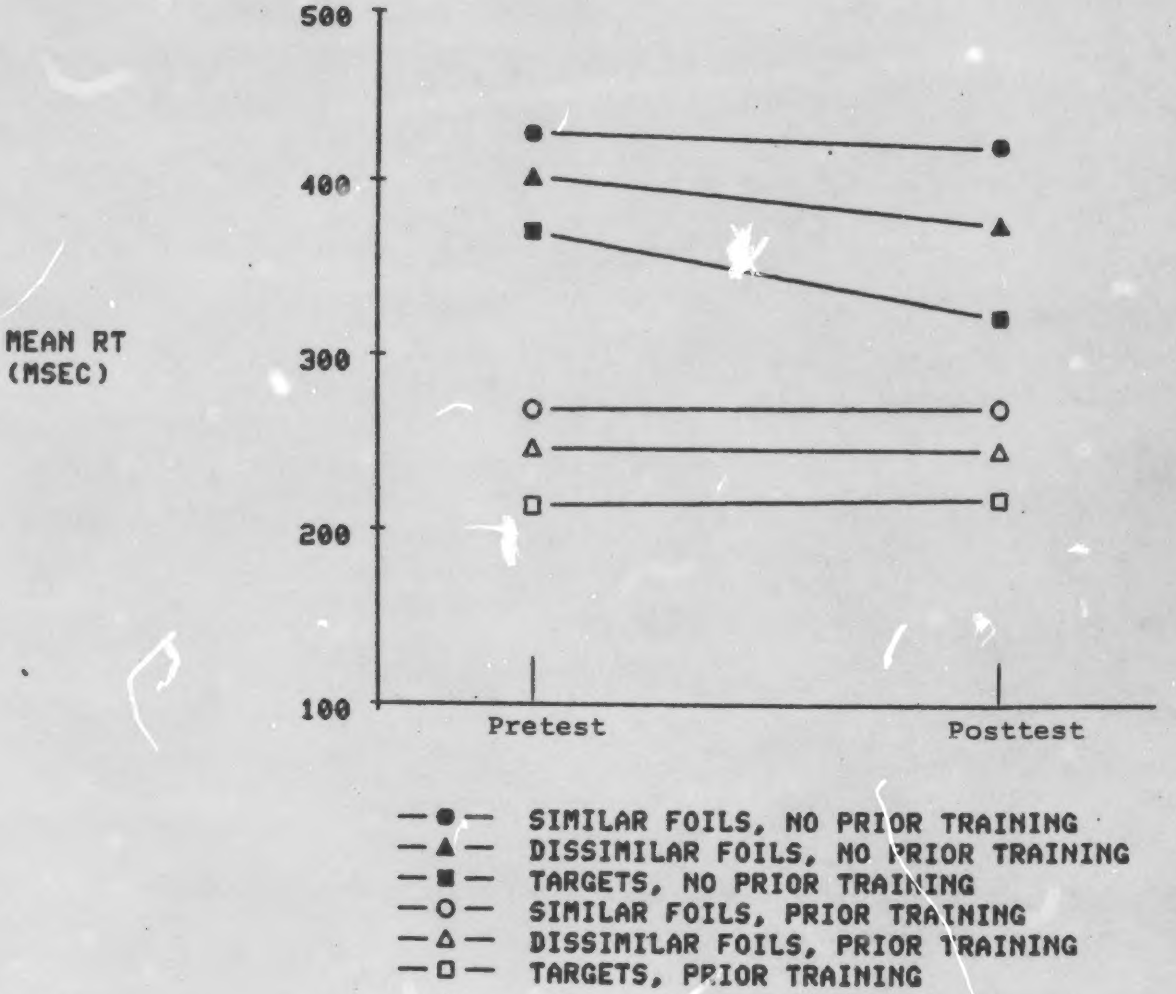


Figure 3.24. Effects of Racer training on mean RT for unit detections. Mean RTs are for correct responses. Subjects in the prior training group completed training on Speed before beginning Racer training.

prior training), with $F_{2,8}=13.1$, $p=.003$ and $F_{1,4}=10.6$, $p=.03$, respectively. Our conclusion is that effective training systems must provide feedback that is explicitly correlated with the skilled performance that is to be acquired.

Span of apprehension. Transfer of Racer training to the visual span task occurred for some but not all subjects. For example, of the six subjects, three showed significant increases in measures of visual span--the average length in letter spaces of text they could report when reading a briefly exposed phrase. The pre- and posttest spans for these subjects were: 15.0 - 17.2 (KG; $F_{1,156}=7.7$, $p<.01$), 11.7 - 14.2 (SZ; $F_{1,156}=8.8$, $p<.005$), and 8.9 - 11.2 (OD; $F_{1,156}=14.6$, $p<.001$). But two subjects showed decreases in visual span: 8.9 - 5.7 (JS; $F_{1,152}=23.5$, $p<.001$) and 15.9 - 13.8 (MG; $F_{1,156}=3.3$, $p<.005$). The three subjects showing increases in span were also those who achieved the highest levels of efficiency in decoding words at the end of their Racer training.

While there were individual differences in the effects of training on the measure of span, in general subjects showed a reduction in RT for reporting words and phrases in the span task. The mean RTs are shown in Figure 3.25. The main effect of training approached significance, with $F_{1,4}=6.46$, $p=.06$. Results of analyses for individual subjects showed significant training effects on latencies for four subjects, and nonsignificant

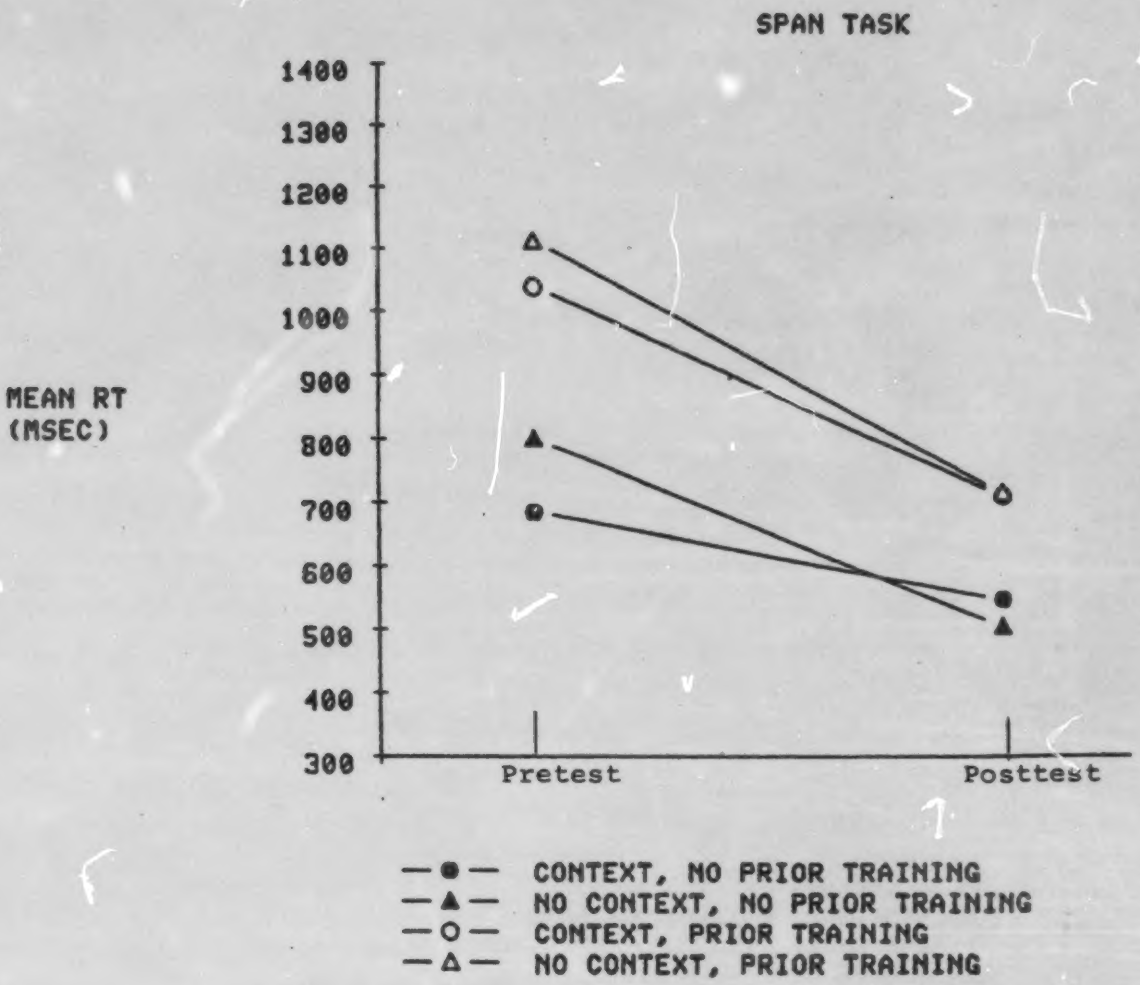


Figure 3.25. Effects of Racer training on the mean RT in the span of apprehension task. Subjects in the prior training group completed training on Speed before beginning Racer training.

reductions in latency for the remaining two subjects. Two of the subjects (JS and MG) who showed significant decreases in RT of 442 msec (JS) and 203 msec (MG) also showed decreases in visual span following training, suggesting a trade-off of speed for accuracy.

Inference task. There were no general effects of Racer training on speed or accuracy of performance on the inference task; nor were there any significant changes in reading rate. In the individual analyses, there were no significant changes in the number of connectives correctly identified or in reading rate. However, three subjects showed significant reductions in RT for judging appropriateness of connectives. Pre- and posttest RTs for these subjects were: 2586 - 2071 msec (JS, $t_{17}=2.00$, $p<.05$), 2366 - 1644 msec, (SZ, $t_{29}=7.20$, $p<.001$) and 5136 - 3315 msec (OD; $t_{26}=3.25$, $p<.005$).

Summary. The effectiveness of the Racer system for improving subjects' efficiency in decoding words was clearly demonstrated for every subject in our sample. While training focused on improving speed of responding with controls for accuracy, there were gains of about 15% in accuracy of pronunciation, with final accuracy levels in pronouncing words averaging 84%. Transfer of word decoding skill to a pseudoword pronunciation task indicated that subjects had acquired an ability to phonologically decode orthographic patterns of English as well as to efficiently identify words. There was no evidence of transfer of training to a perceptual subskill involved in decoding as indexed by the unit detection task. Transfer of training to general measures of reading skill, namely the visual span and inferential processing tasks, was not a general finding although instances of transfer were found for individual subjects.

3.4 Ski-Jump Training

Individual training records for Ski-Jump training are presented in Figures 3.26 through 3.39 (a more detailed account of training is given by Gillotte, 1983). The first four subjects (JS, KG, MG, OD) received prior training using the Speed and Racer systems while the final three subjects (EK, LN, AC) did not. The even-numbered figures (3.26, 3.28,...,3.38) present mean RT for each training session for responses that were correct. Recall that the RT includes the total time from the initial flash of the target word (presented with pre-and post-exposure masks) until the subject responds. Successive flashes of the masked stimulus occur during that interval, with flash durations starting at 18 msec (1 "paint" of the screen) and incremented by that amount on each succeeding flash. Stimulus onset asynchronies were initially 2000 msec., and were decreased during the course of training on a schedule determined for each subject on the basis of his or her performance. Subjects began

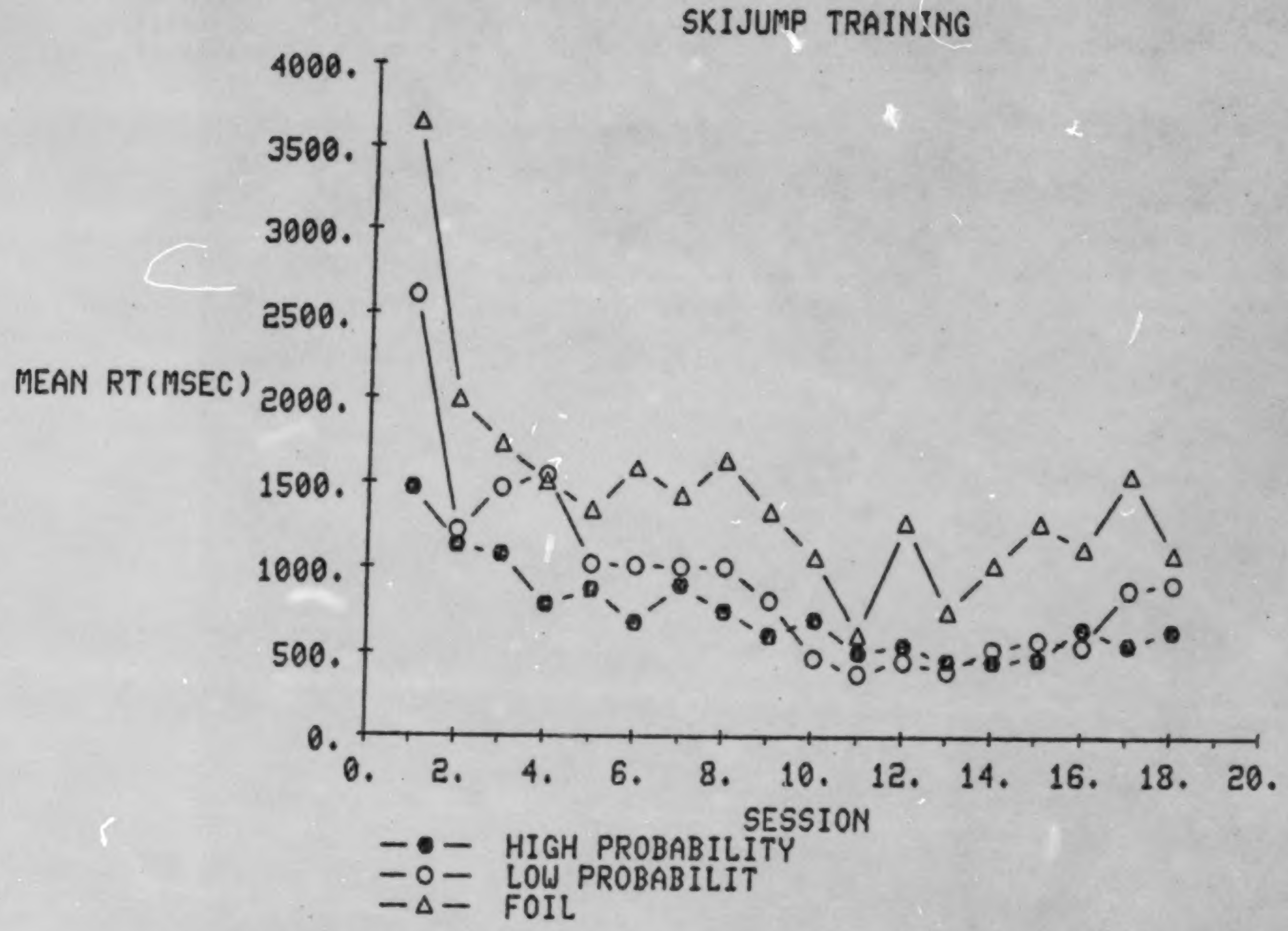


Figure 3.26. Ski-Jump training records for subject JS. Plotted are the mean RTs for each training session for (a) high probability target words presented as the last or next-to-last test item, (b) low probability target words presented as the first test item, and (c) foils or words which are unrelated to the context. Session 1-8 employed high constraining contexts, and the remaining sessions, low constraining contexts.

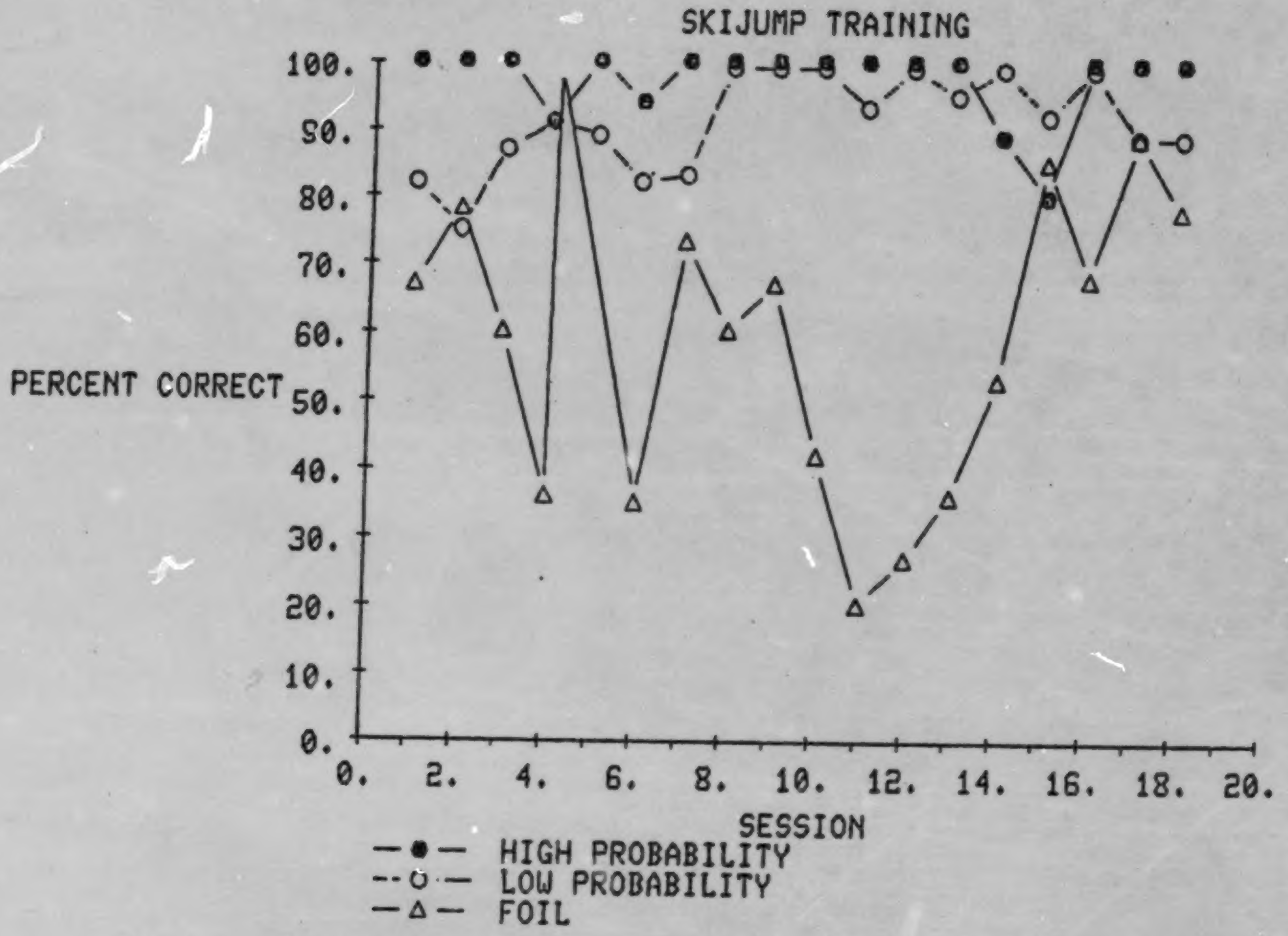


Figure 3.27. Ski-Jump training records for subject JS. Plotted is the mean percent of correct semantic appropriateness judgements for high probability targets, low probability targets, and foils. Sessions 1-8 employed high constraining contexts, and the remaining sessions, low constraining contexts.

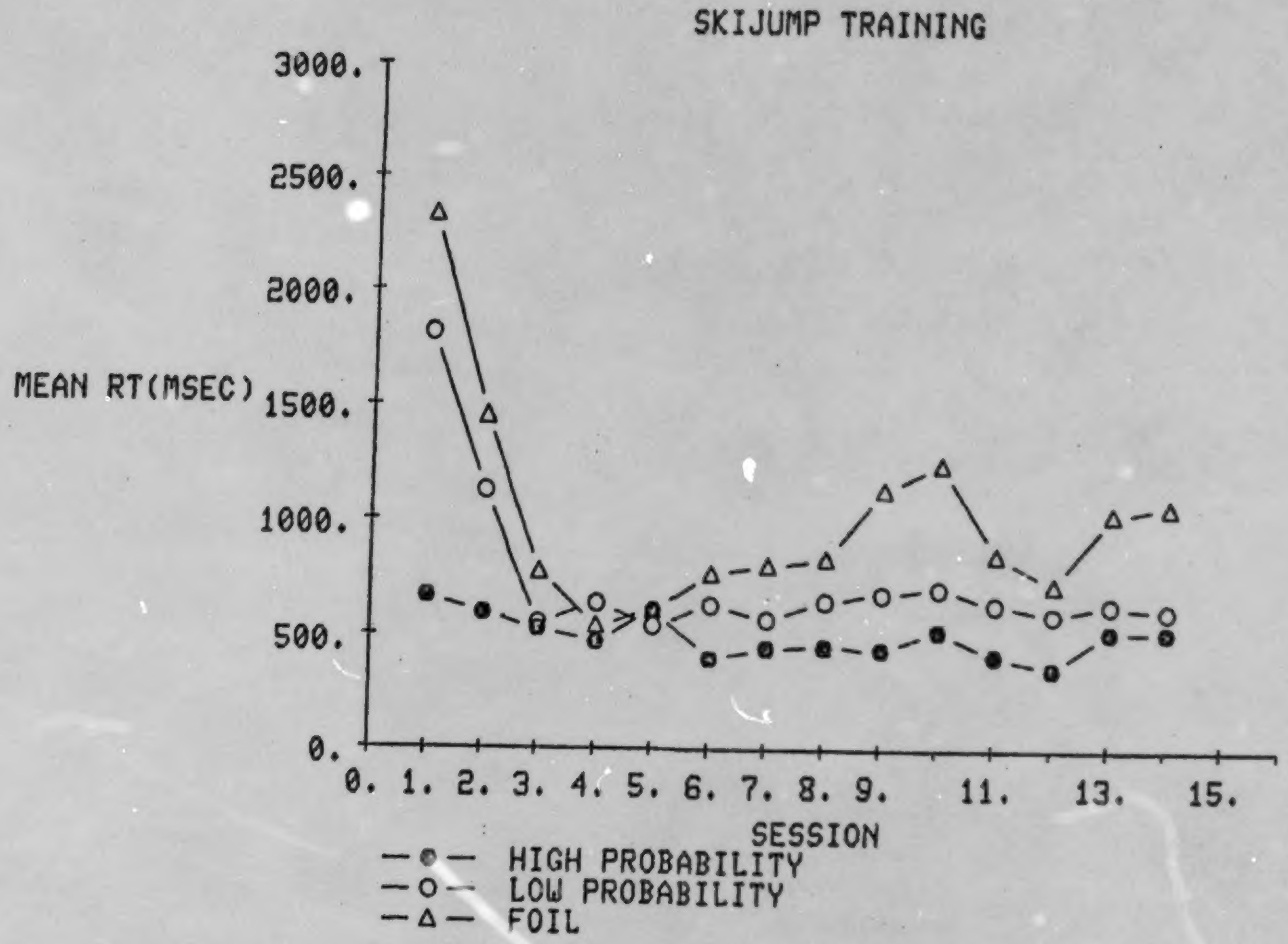


Figure 3.28. Ski-Jump training records for subject KG. Plotted are the mean RTs for each training session for (a) high probability target words presented as the last or next-to-last item, (b) low probability target words presented as the first test item, and (c) foils or words which are unrelated to the context. Sessions 1-7 employed high constraining contexts, and the remaining sessions, low constraining contexts.

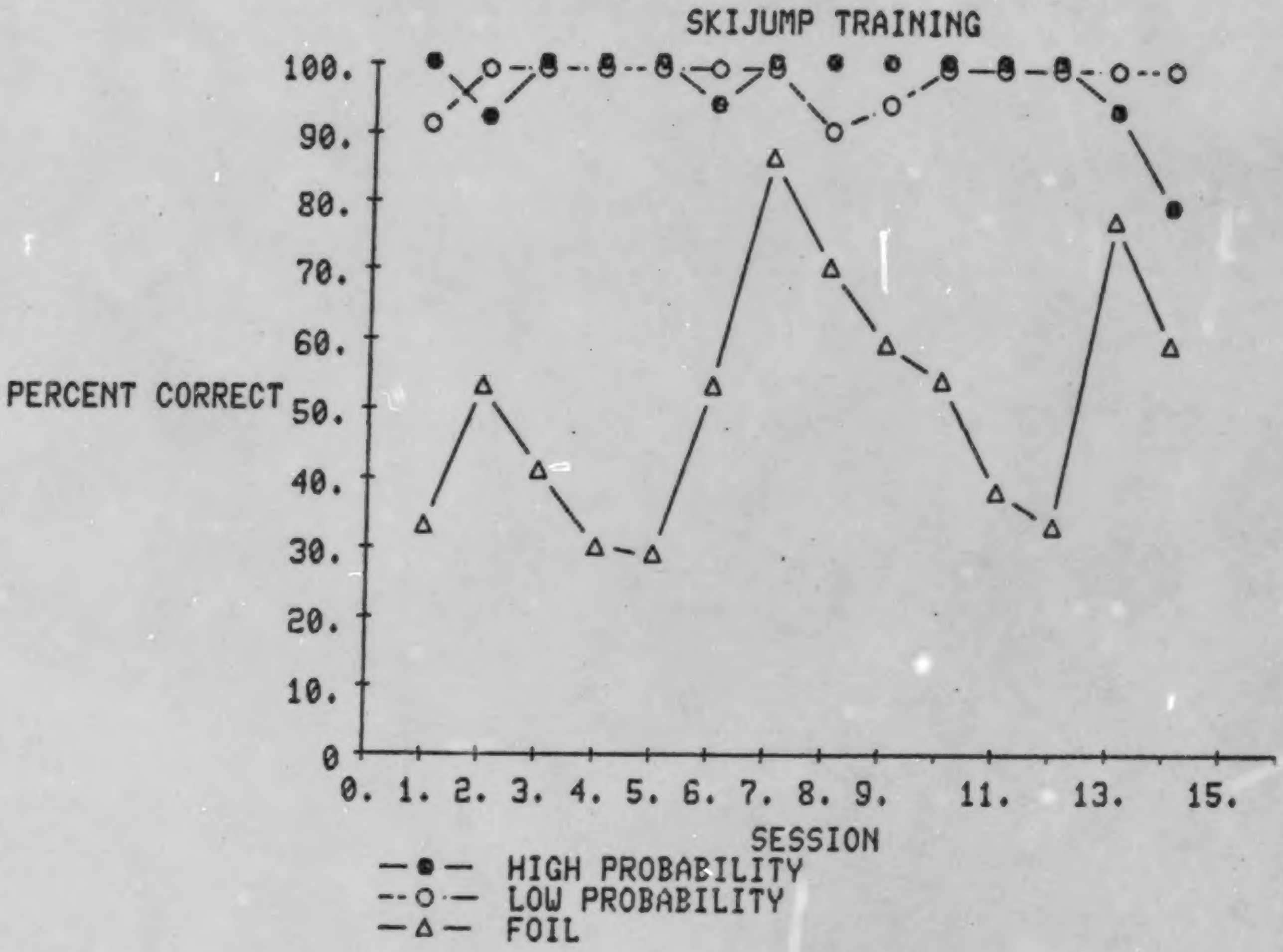


Figure 3.29. Ski-Jump training records for subject KG. Plotted is the mean percent of correct semantic appropriateness judgments for high probability targets, low probability targets, and foils. Sessions 1-7 employed high constraining contexts, and the remaining sessions, low constraining contexts.

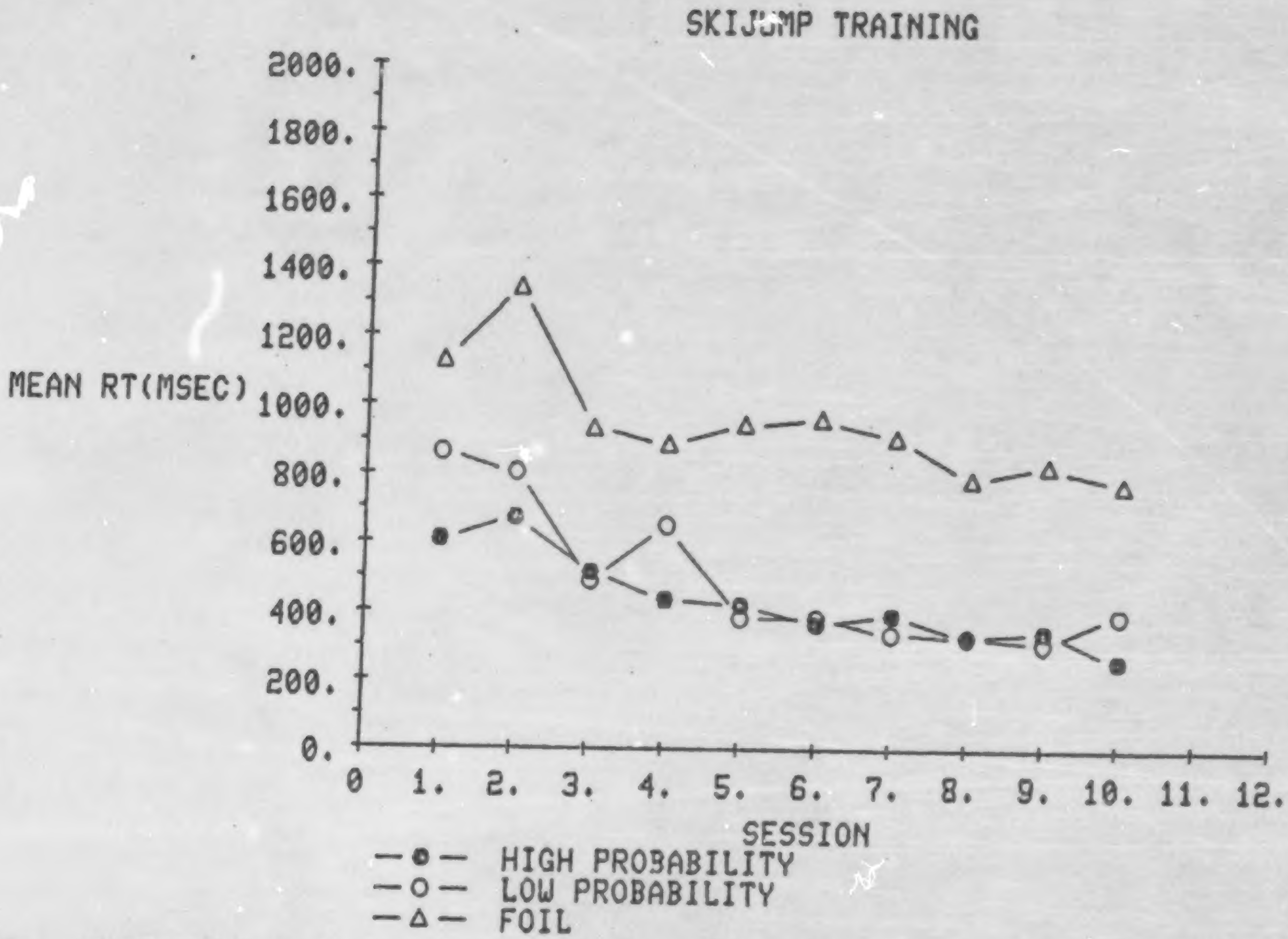


Figure 3.30. Ski-Jump training records for subject MG. Plotted are the mean RTs for each training session for (a) high probability target words presented as the last or next-to-last item, (b) low probability target words presented as the first test item, and (c) foils or words which are unrelated to the context. Sessions 1-6 employed high constraining contexts, and the remaining sessions, low constraining contexts.

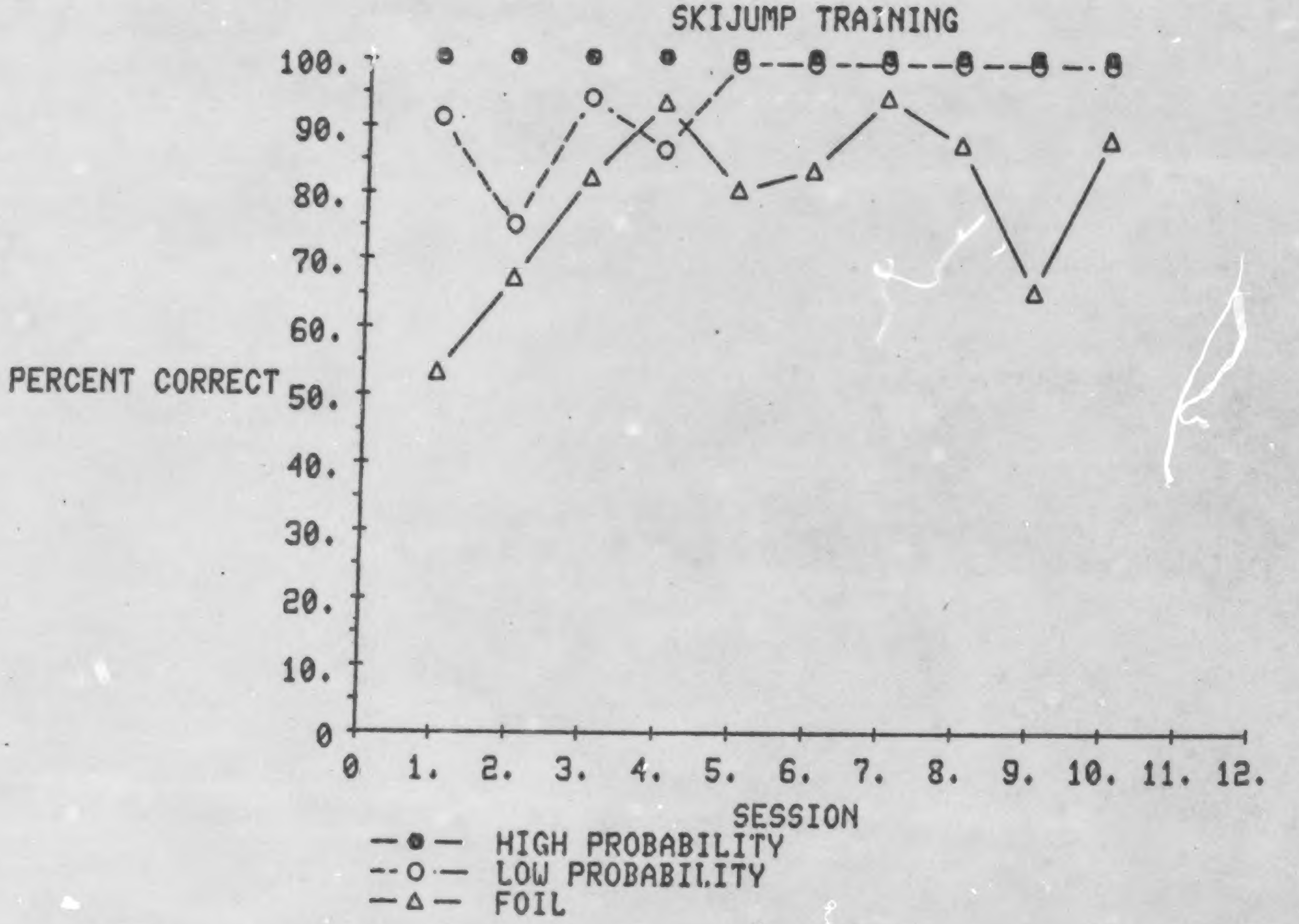


Figure 3.31. Ski-Jump training records for subject MG. Plotted is the mean percent of correct semantic appropriateness judgments for high probability targets, low probability targets and foils. Sessions 1-6 employed high constraining contexts, and the remaining sessions, low constraining contexts.

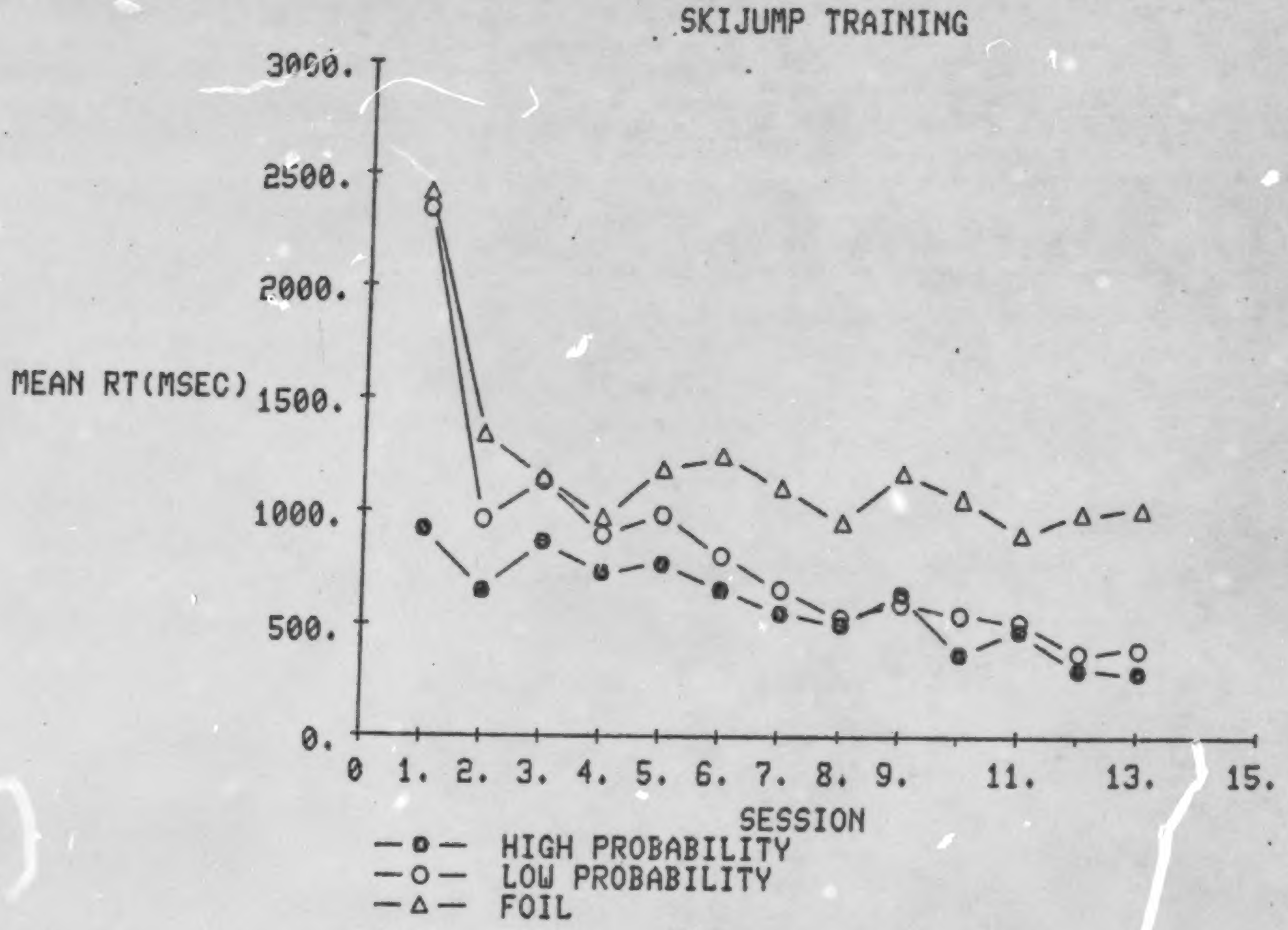


Figure 3.32. Ski-Jump training records for subject OD. Plotted are the mean RTs for each training session for (a) high probability target words presented as the last or next-to-last test item, (b) low probability target words presented as the first test item, and (c) foils or words which are unrelated to the context. Sessions 1-8 employed high constraining contexts, and the remaining sessions low constraining contexts.

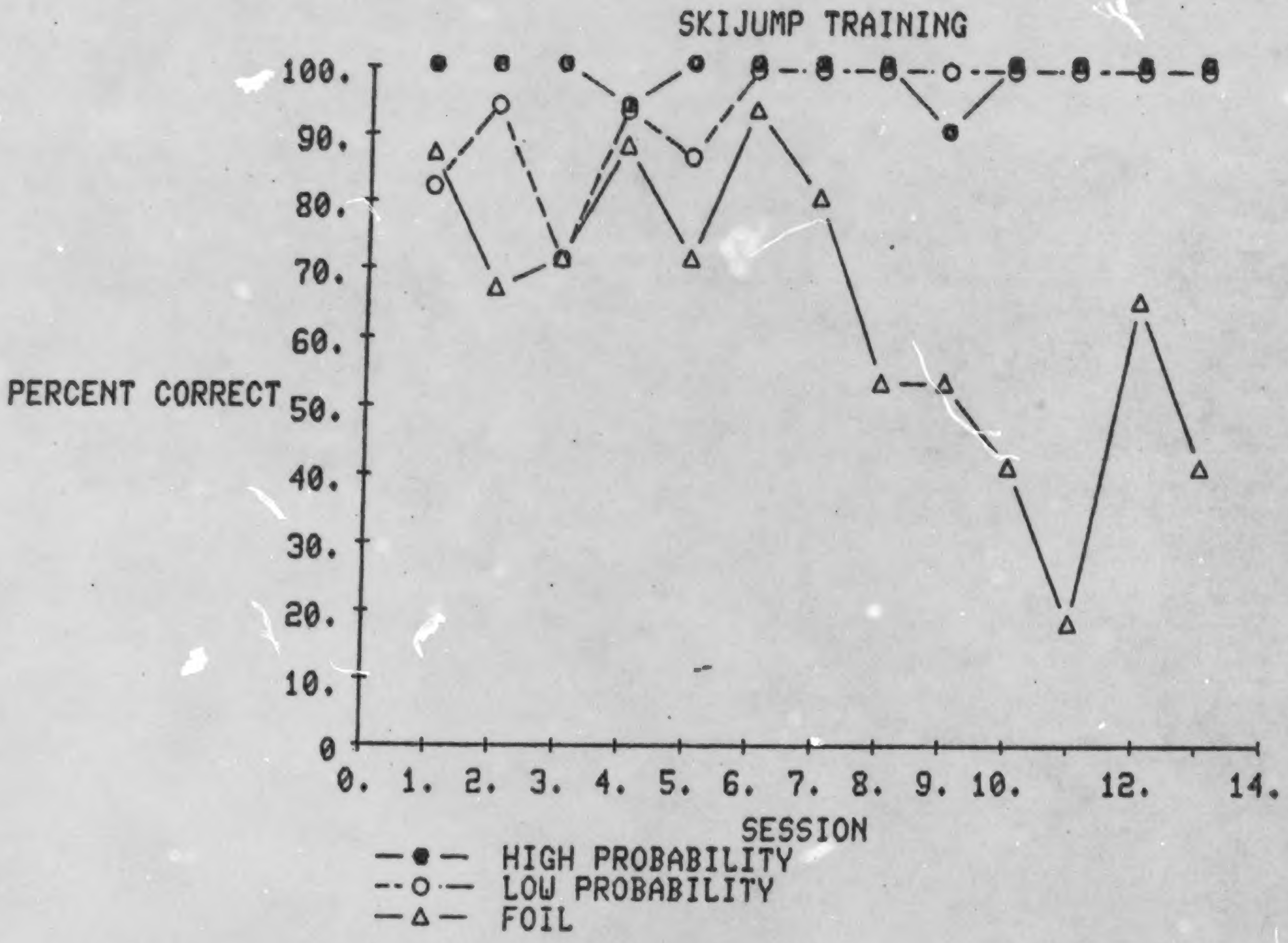


Figure 3.33. Ski-Jump training records for subject OD. Plotted is the mean percent of correct semantic appropriateness judgments for high probability targets, low probability targets, and foils. Sessions 1-8 employed high constraining contexts, and the remaining sessions low constraining contexts.

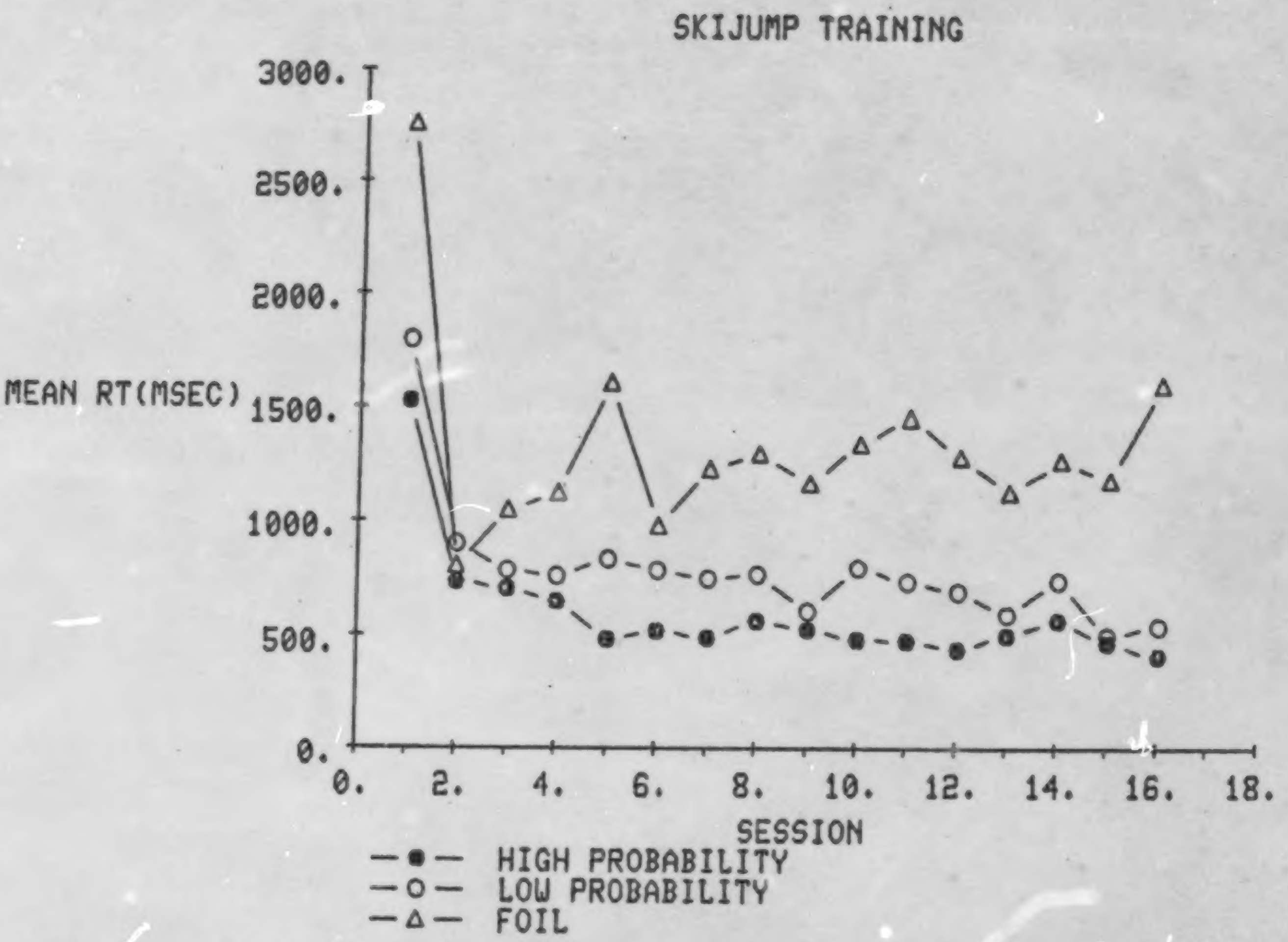


Figure 3.34. Ski-Jump training records for subject EK. Plotted are the mean RTs for each training session for (a) high probability target words presented as the last or next-to-last test item, (b) low probability target words presented as the first test item, and (c) foils or words which are unrelated to the context. Sessions 1-9 employed high constraining contexts, and the remaining sessions, low constraining contexts.

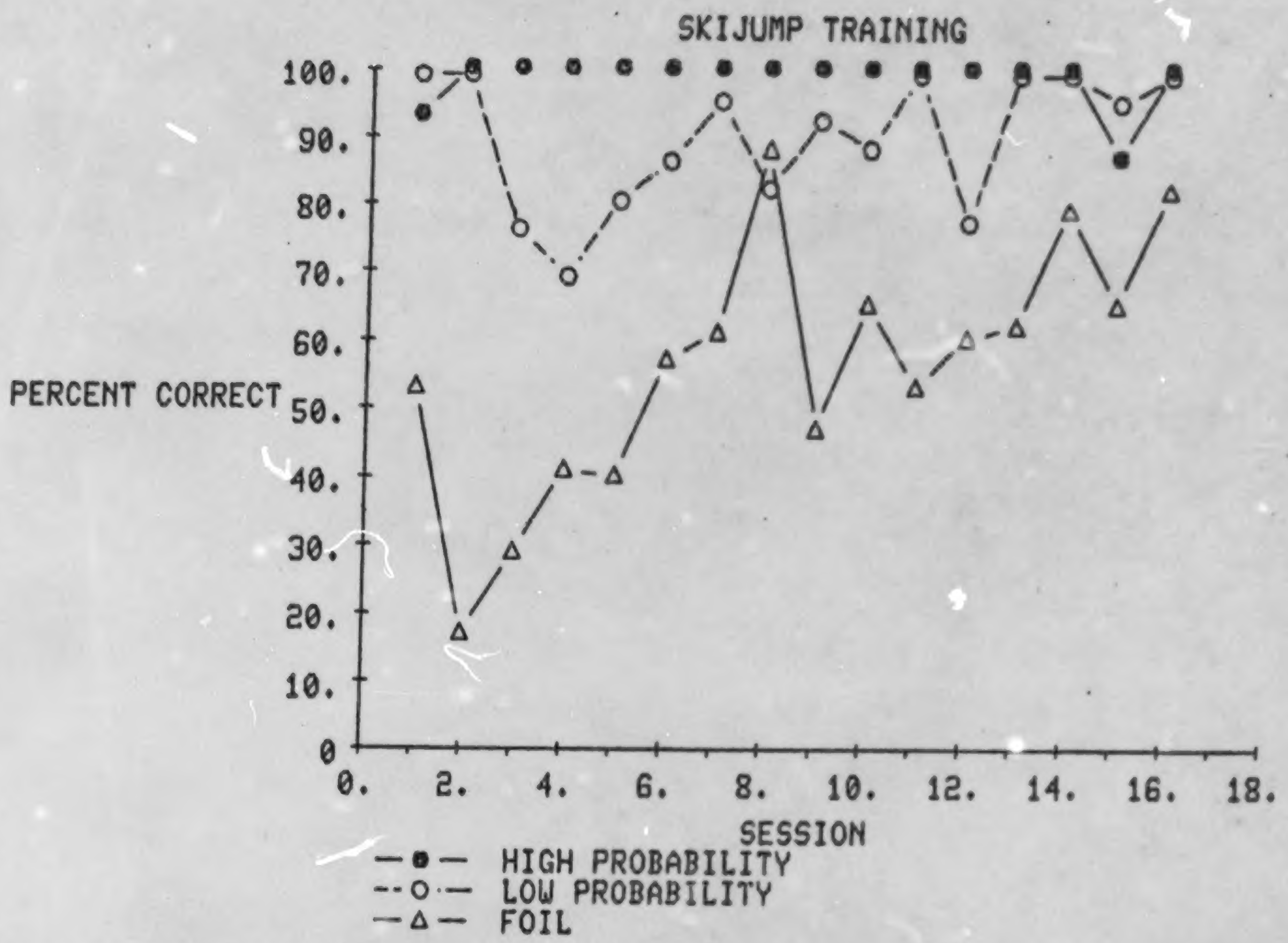


Figure 3.35. Ski-Jump training records for subject EK. Plotted is the mean percent of correct semantic appropriateness judgments for high probability targets, low probability targets, and foils. Sessions 1-9 employed high constraining contexts, and the remaining sessions, low constraining contexts.

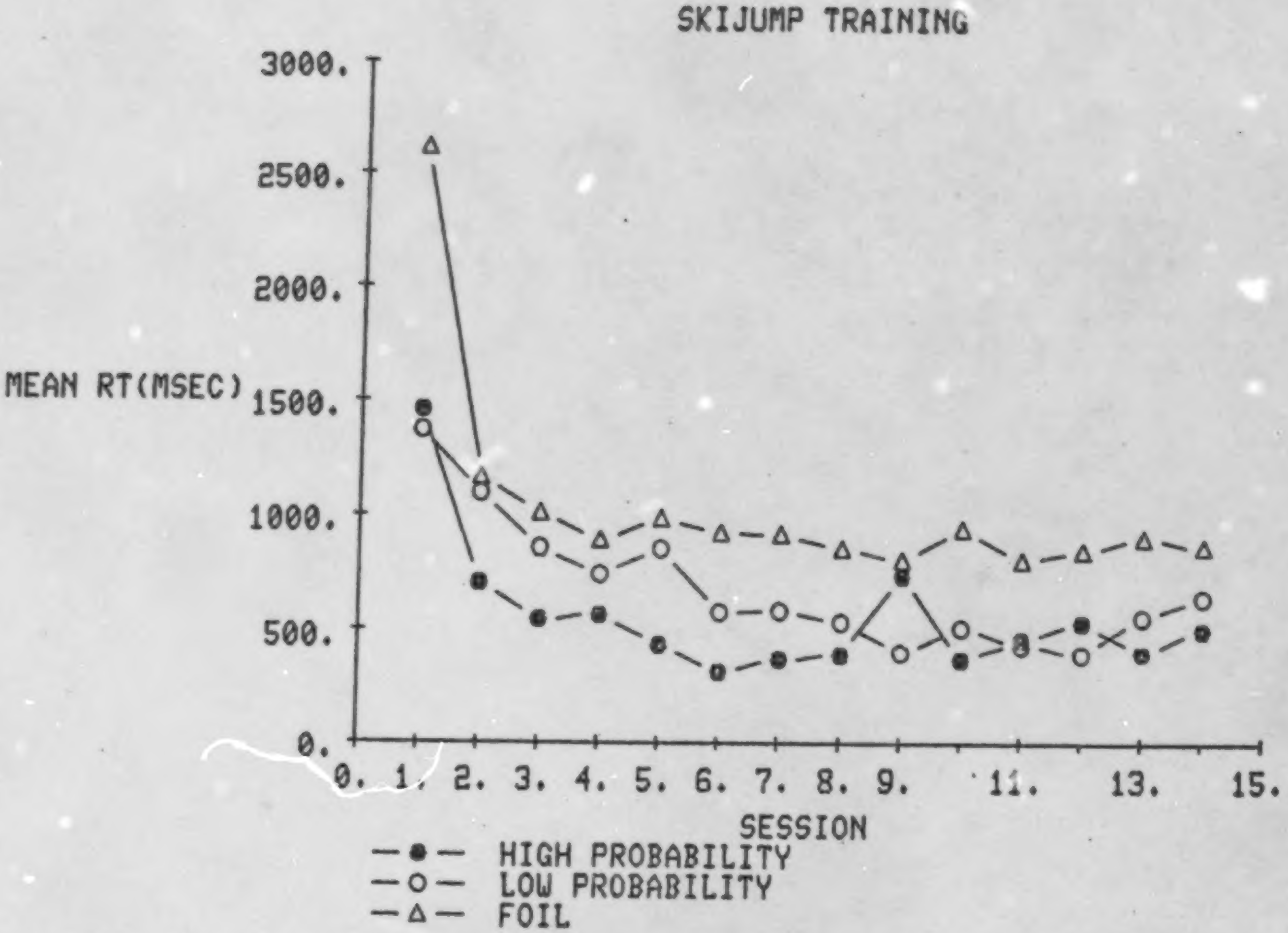


Figure 3.36. Ski-Jump training records for subject LN. Plotted are the mean RTs for each training session for (a) high probability target words presented as the last or next-to-last test item, (b) low probability target words presented as the first test item, and (c) foils or words which are unrelated to the context. Sessions 1-8 employed high constraining contexts, and the remaining sessions, low constraining contexts.

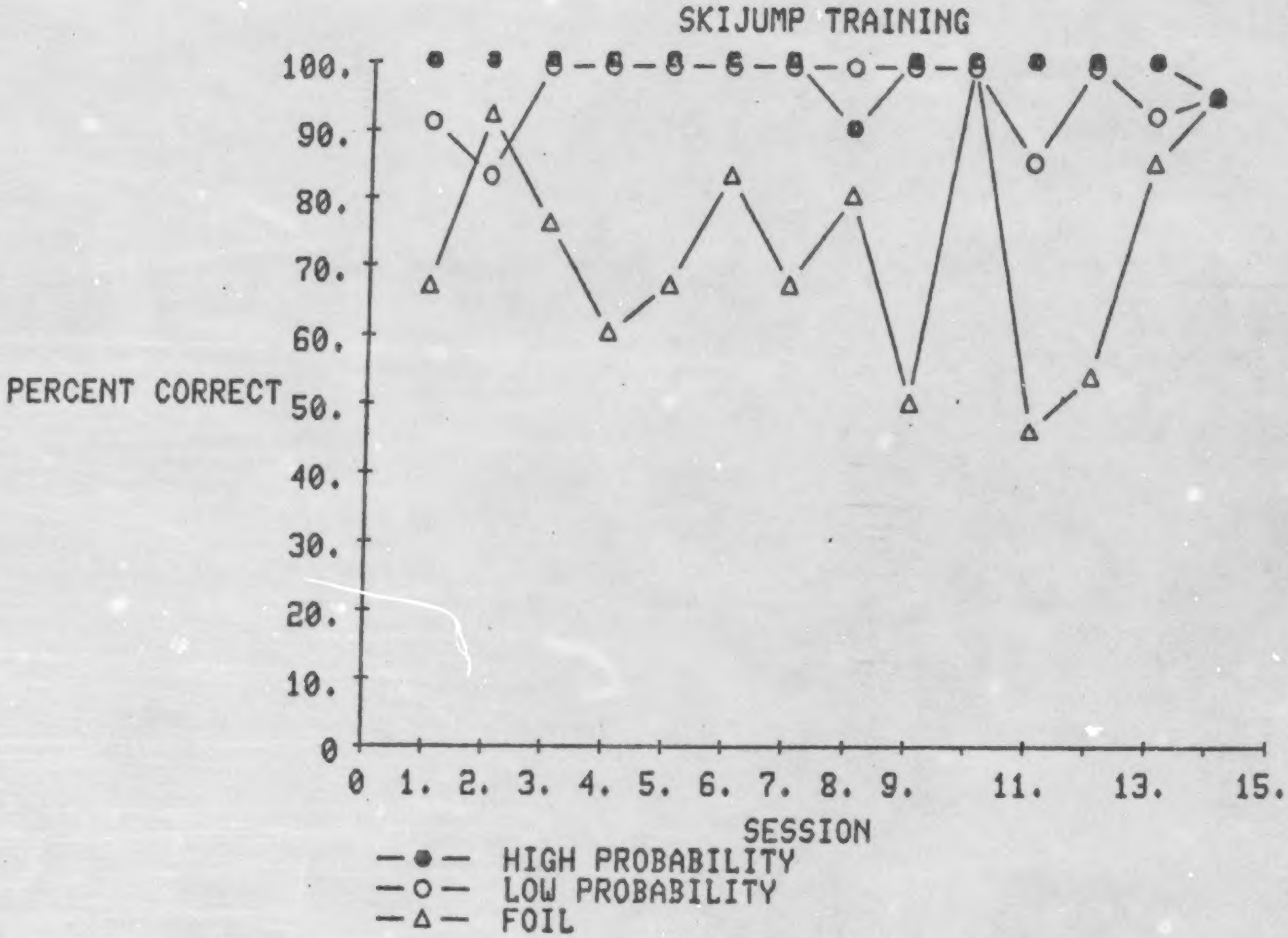


Figure 3.37. Ski-Jump training records for subject LN. Plotted is the mean percent of correct semantic appropriateness judgments for high probability targets, low probability targets and foils. Sessions 1-8 employed high constraining contexts, and the remaining sessions, low constraining contexts.

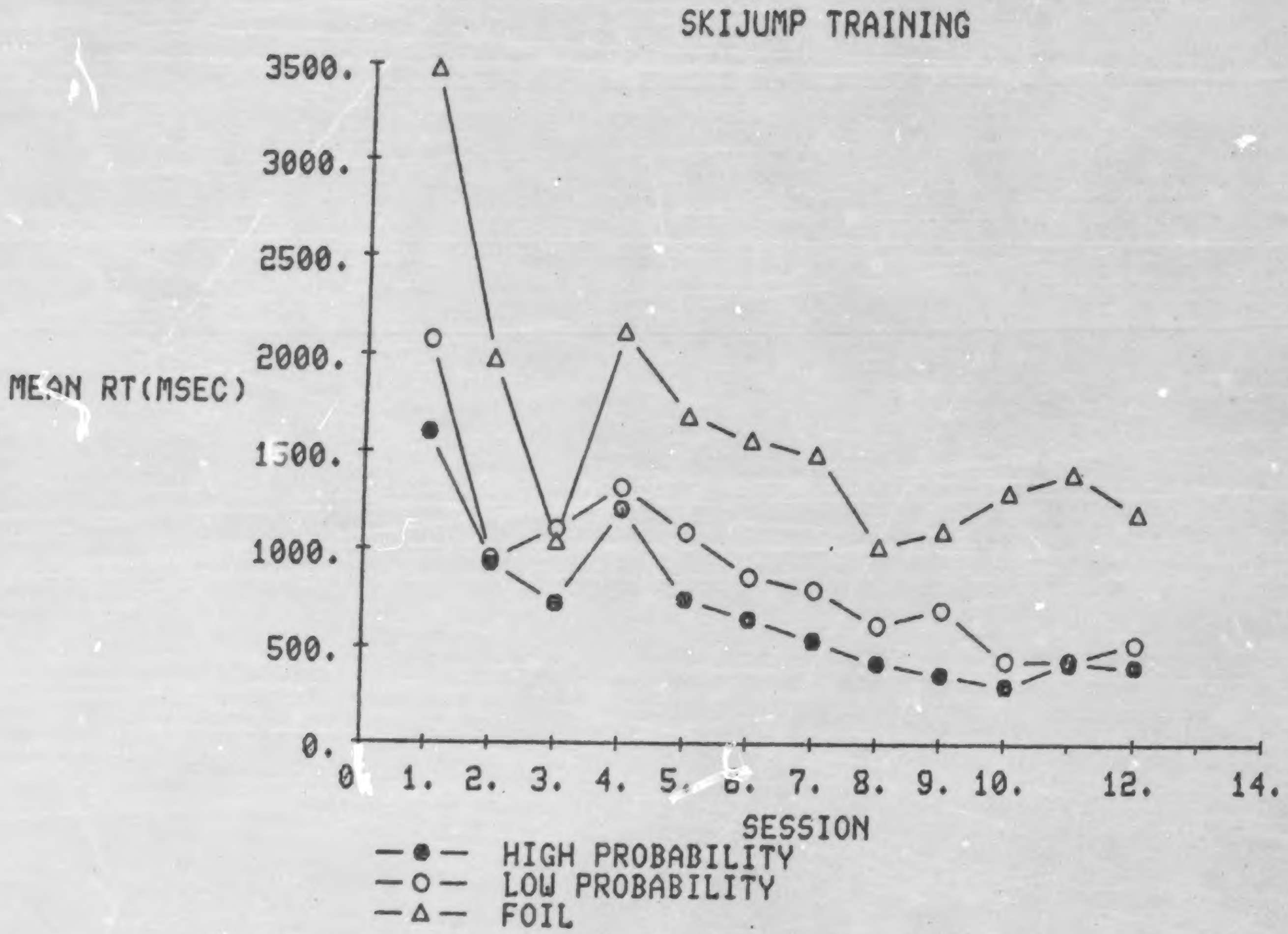


Figure 3.38. Ski-Jump training records for subject AC. Plotted are the mean RTs for each training session for (a) high probability target words presented as the last or next-to-last test item, (b) low probability target words presented as the first test item, and (c) foils or words which are unrelated to the context. Sessions 1-6 employed high constraining contexts, and the remaining sessions, low constraining contexts.

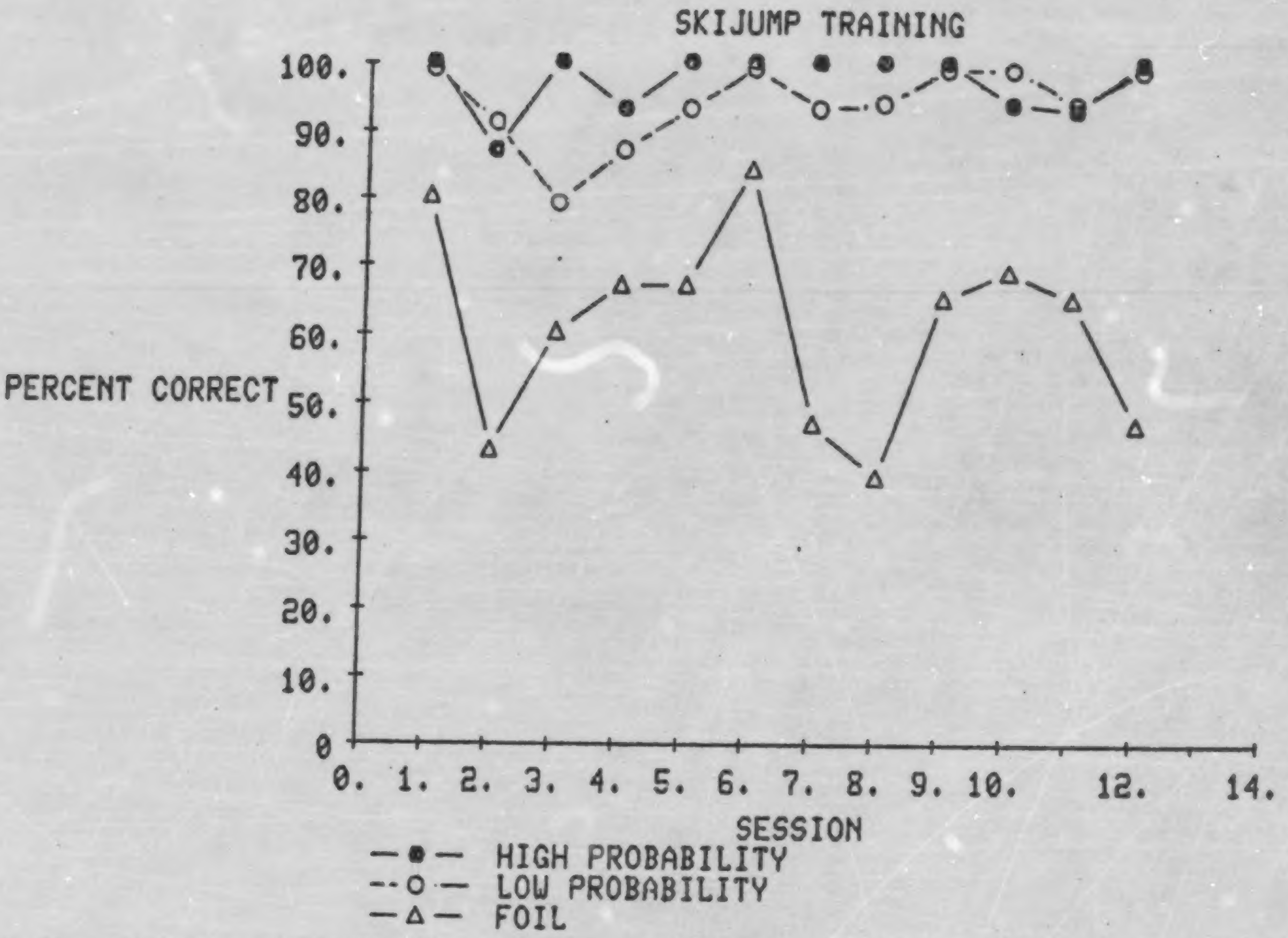


Figure 3.39. Ski-Jump training records for subject AC. Plotted is the mean percent of correct semantic appropriateness judgments for high probability targets, low probabilty targets, and foils. Sessions 1-6 employed high constraining contexts, and the remaining session, low constraining contexts.

training with context sentences that were classified as highly constraining for the semantic domain of the target words. Approximately half way through training, they were switched to more difficult, less constraining sentences. The three dependent variables plotted are mean RTs for foils, mean RTs for high probability target words (which occurred last in the set of test words for a given sentence), and, most important, mean RTs for low probability targets occurring as the first test item (the one immediately following the presentation of the context sentence). Finally, the odd-numbered figures (3.27, 3.29,...3.39) present, for each subject, the mean percent of correct judgments of semantic acceptability for the high and low probability targets and foils in each training session.

Some general features of these training records are noteworthy. At the beginning of training, subjects were much faster in judging high probability target words than they were low probability targets presented immediately after the context sentence. Performance for these low probability targets is close to that for foils -- words unrelated to the context and, by definition, incapable of contextual priming. Performance on the high probability targets, which were presented at the end of a series of target items, reflects the degree of priming possible for a subject after viewing not only the context sentence, but also a series of exemplars of the constrained semantic domain. Successful training should result in improvement in performance on the immediately-presented, low probability targets to the point that RT for those items will approach the RT for the high probability targets. The performance of our subjects in the final session of practice conforms to this pattern. At the end of training, RTs for high and low probability targets are nearly the same, and are substantially faster than those for foils. The mean RT (averaged across seven subjects) at the end of training was 430 msec for high probability targets, 572 msec for low probability targets, and 1083 msec for foils.

In addition, the training records show an initial large decrease in mean RT for foils as well as targets, followed by a period of more gradual but steady improvement in performance in which RT for low probability targets continues to decline while that for foils reaches a plateau. The final levels of performance attained by our subjects represent extremely fast reaction times, particularly since the task involves a visually masked stimulus as well as a judgment of semantic appropriateness which itself involves recognition of the target word.

While all subjects maintained high levels of accuracy for target words, there was greater variability in their accuracy in rejecting foils. Two subjects, OD (Figure 3.33) and AC (Figure 3.39) appear to have accepted a high number of errors in order to further reduce RTs in the last 5-6 training sessions. The other subjects maintained consistent levels of accuracy over those training sessions.

Transfer of Training

Context priming task. Skill improvements resulting from training with the Ski-Jump game were assessed using a context-priming task similar in form to the Ski-Jump training task. Subjects read a context sentence and were then presented target words in a series of exposures beginning at 18 msec and incremented by 18 msec on each subsequent exposure. Each exposure was preceded and followed by a masking stimulus. The subject's task was to judge whether or not the target word fit the preceding sentence context. The subjects in the analysis of variance included four who had prior training on Speed and Racer, and three with no prior training. The results are shown in Figure 3.40. Subjects in both groups showed large decreases in RT, reaching posttest levels of 600-900 msec. The main effect of training was significant with $F_{1,5}=22.3$, $p=.005$. The improvement in performance included low probability target words as well as high probability words. Finally, there was a significant 3-way interaction of group, training, and context ($F_{1,5}=7.60$, $p=.04$). Subjects who had not received prior training in perceptual and word decoding skills showed a higher initial mean RT on the Ski-Jump criterion task, and a larger benefit of training, reaching final levels of performance similar to those of the previously trained subjects. Each of the context sentences employed in the criterion task involved a unique semantic domain, as had the materials used in the Ski-Jump training task itself. Thus, skill

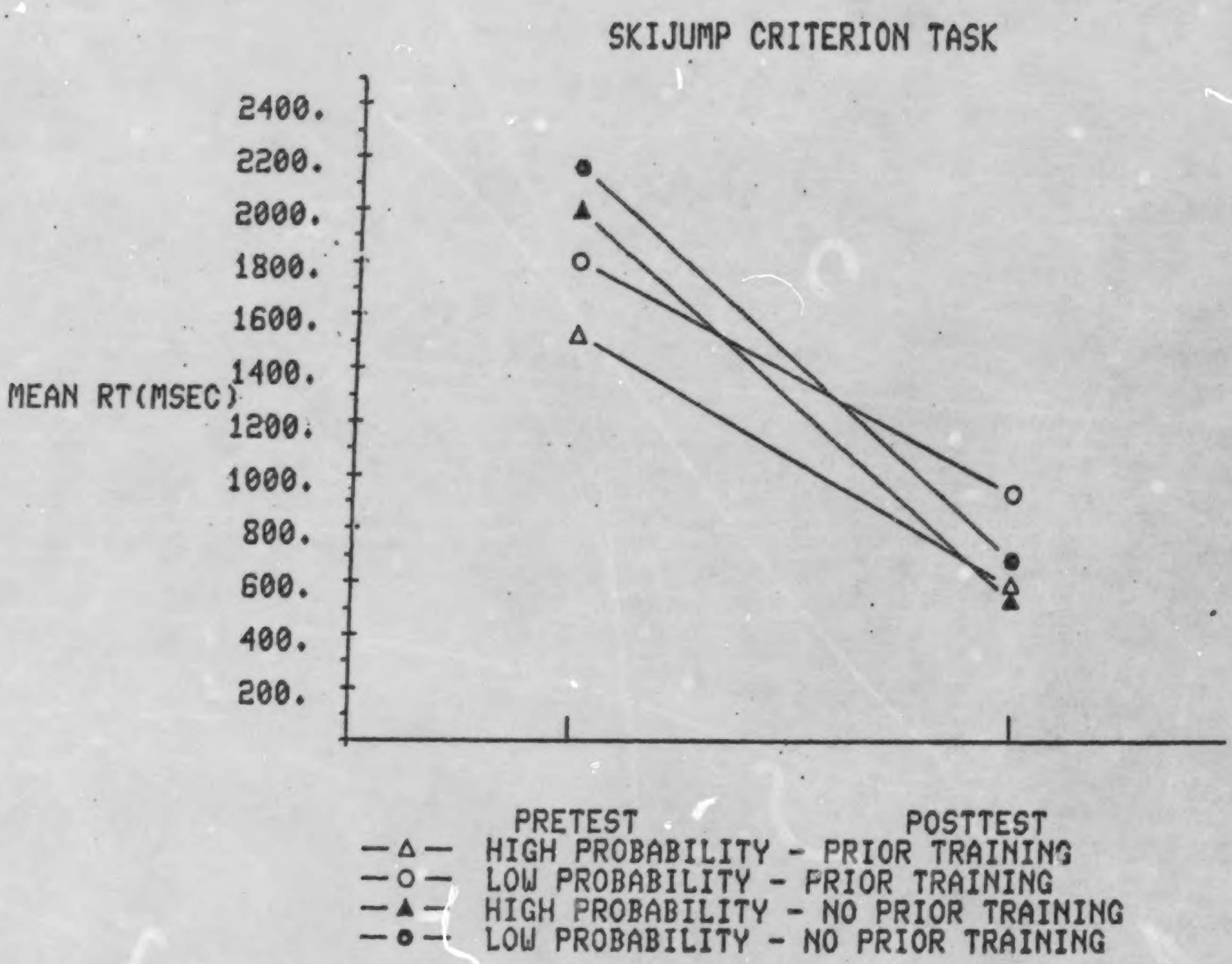


Figure 3.40. Results of Ski-Jump training as measured by performance on the criterion context priming task. Subjects in the prior training group had completed Speed and Racer prior to undertaking Ski-Jump.

improvement cannot be attributed to knowledge of the semantic domains of the sentences themselves. Rather, it can be attributed to the development of a general facility for establishing and using frame-based activations of concepts in semantic memory. Training effects may also be attributable to improvements in ability to judge the semantic appropriateness of targets, but our results argue against this as the primary source of the effects. While the semantic decisions themselves were designed to be easy in that fine distinctions in meaning were not required, the most difficult judgments of semantic acceptability remained those for foils. Indeed, the effects of training on RT for foils were smaller than those for targets and were more restricted to the early training sessions.

Reference group experiment. The magnitude of the skill improvements on the Ski-Jump criterion task can be demonstrated by comparing the final performance levels of our trainees with levels achieved by four reference groups of untrained subjects who differed greatly in their reading abilities. Gillotte (1983) compared the performance of four groups of readers representing the following percentile ranges on the Nelson Denny Reading Test: (1) 2-17, (2) 22-42, (3) 71-80, and (4) 90-99. The results are shown in Figure 3.41. There was a significant interaction of readers' ability level and target probability ($F_{3,11}=4.49$, $p=.027$), and a significant main effect of target probability ($F_{1,11}=13.9$, $p=.003$). Comparing our trainees with these

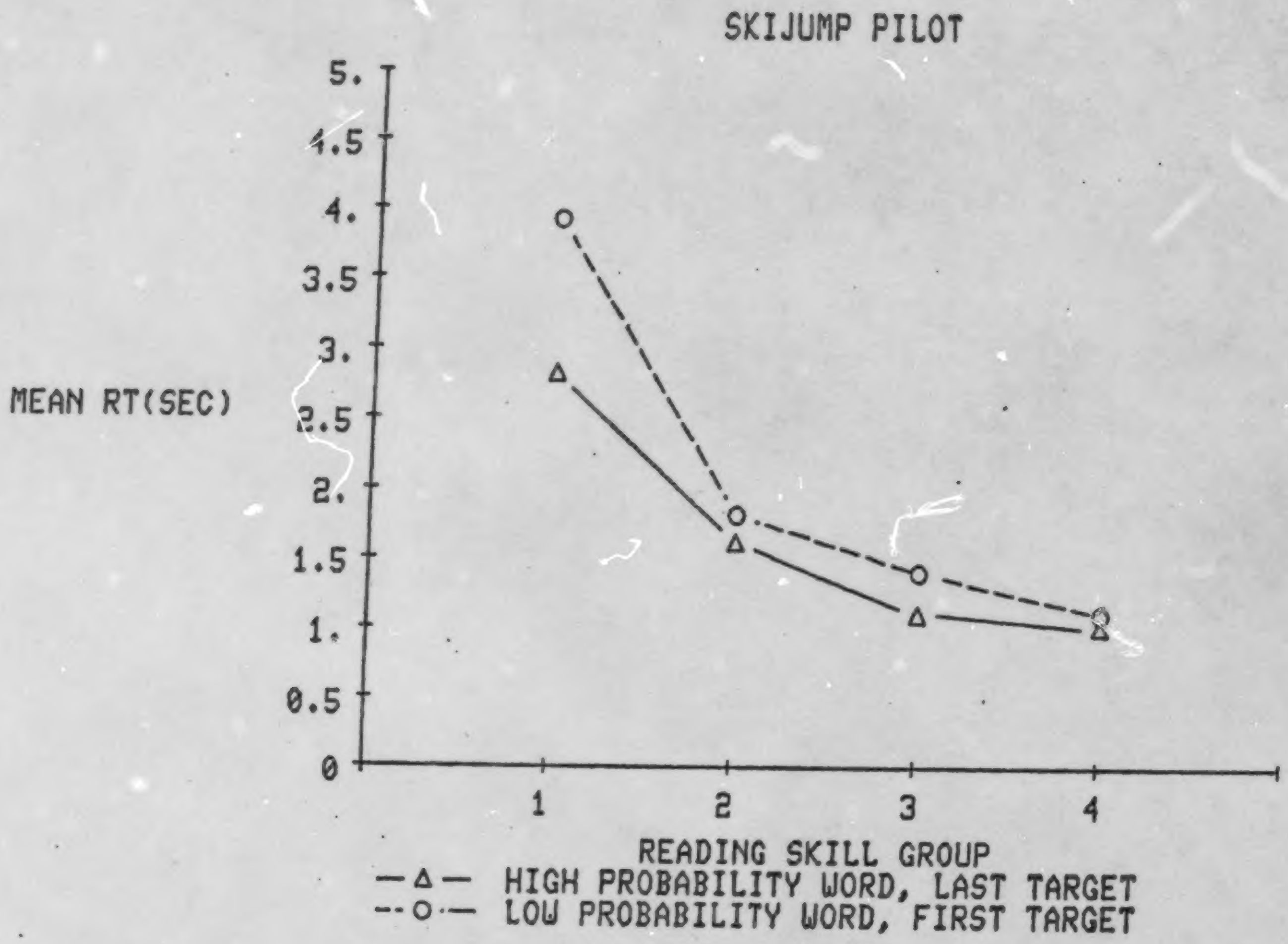


Figure 3.41. Performance levels of four reference groups of subjects on the context priming task.

reference groups, we see that their initial mean RTs were similar to those for group 2, and their final levels of performance exceeded those of the highest ability readers.

Unit detection task. The effect of Ski-Jump training on the development of perceptual subskills of reading was evaluated using the unit detection task. Mean unit detection latencies are shown in Figure 3.42. Again, there were two groups of subjects: those who had prior training with Speed and Racer, and those had no prior training. There was a significant difference in mean RT for the two groups ($F_{1,5}=14.6$, $p=.01$). Subjects who had completed their Speed training several months before continued to perform at asymptotic levels on the unit detection task with a mean RT of 218 msec. The corresponding mean RT for subjects who had not been previously trained was 486 msec. While the main effect of Ski-Jump training on unit detection was not significant, there was a significant interaction between subject group and Ski-Jump training ($F_{1,5}=12.2$, $p=.017$). While there was a 105 msec decrease in mean RT for targets for the group having no prior training, there was no change in performance for the trained group which was already performing at an asymptotic level. Analyses of performance for individual subjects were consistent with the group results reported above. There were significant training effects for all three subjects who had not received prior Speed and Racer training. Their pre- and posttest RTs on correct trials for words that contained the target unit

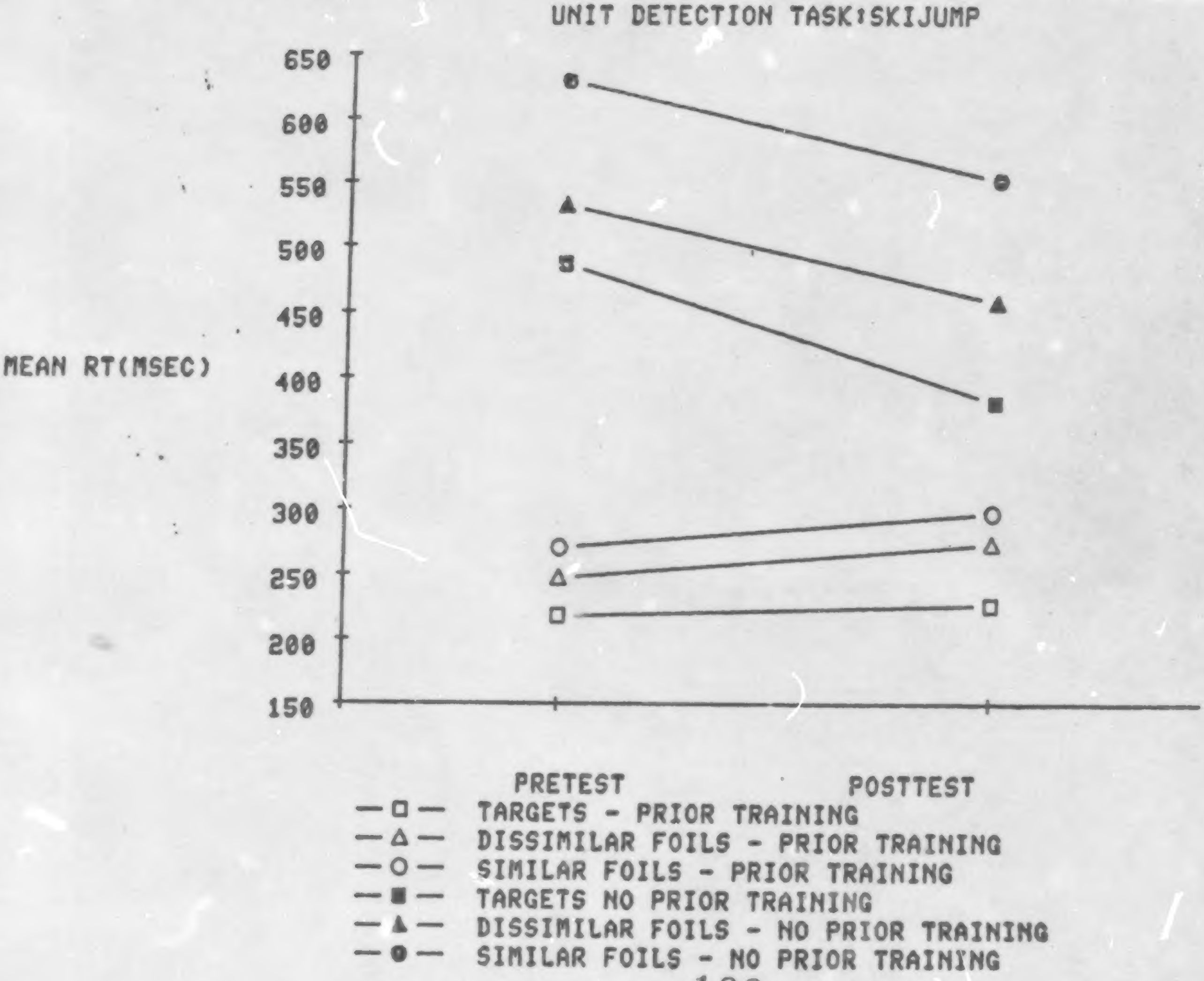


Figure 3.42. Mean RT for unit detections obtained by subjects before and after training on Ski-Jump. Subjects in the prior training group had completed Speed and Racer prior to undertaking Ski-Jump.

were 381-347 msec (EK; $F_{1,123}=5.4$, $p<.05$), 450-312 msec (LN; $F_{1,219}=122.3$, $p<.001$), and 627-485 msec (AC; $F_{1,240}=32.0$, $p<.001$). The first two of these subjects, EK and LN, also showed significant improvement in percent of correct responses; their pre- and posttest scores were 86% to 92% (EK; $F_{1,213}=7.8$, $p<.01$), and 91%-97% (LN; $F_{1,219}=20.9$, $p<.001$). The third subject, AC, showed no change in accuracy, scoring 87% in the pretest and 88% in the posttest. Of the four previously trained subjects, three showed changes in latency for detecting targets of less than 10 msec, and the fourth showed a small (39 msec) increase in detection latency.

A comparison of the unit detection and Ski-Jump tasks offers an explanation for the effects of Ski-Jump training on the development of perceptual skill. Both tasks employ masked stimulus words as stimuli and require rapid encoding of orthographic information for successful task performance. However, while the Ski-Jump game provided some opportunity to develop the perceptual skill, it was not as effective a training environment as the Speed game. It may be, however, that prior training on Speed need not be as thorough or prolonged if training on Ski-Jump is to follow. Subjects who have developed automatic perceptual skills but who have not yet reached an asymptotic level of performance would be expected to continue to develop the perceptual subskill while performing the Ski-Jump task.

Pseudoword and word pronunciation tasks. Effects of Ski-Jump training on the pseudoword pronunciation task are shown in Figure 3.43. There were no improvements in decoding skill for trainees who had no prior training in decoding and who presumably had not developed automatic decoding skills. Additionally, these subjects had longer RTs for two-syllable pseudowords than for one-syllable items. In contrast, trainees who had received prior training using the Racer system continued to show small improvements in speed of pseudoword decoding, and showed no performance differences for one- and two-syllable items before or after training. However, the analysis of variance produced no significant main effect of training, and only a marginal interaction between group, syllable length, and training, with $F_{1,5}=3.57$, $p=.12$. The analyses of variance carried out for individuals produced results that were consistent with the group results. For the group that had no prior training, two subjects showed no change in pseudoword RT and the third showed a small increase in RT. For subjects in the group that received prior training on Speed and Racer: one subject (JS) showed a general reduction in pseudoword RTs, with pre- and posttest RTs of 1057 msec and 820 msec ($F_{1,144}=17.8$, $p<.005$); one (OD) showed reductions in RT for pseudowords derived from high frequency words (from 404 msec to 258 msec) and for 2-syllable pseudowords derived from low frequency words (from 354 to 257 msec) but not for the one-syllable items derived from low frequency words (for

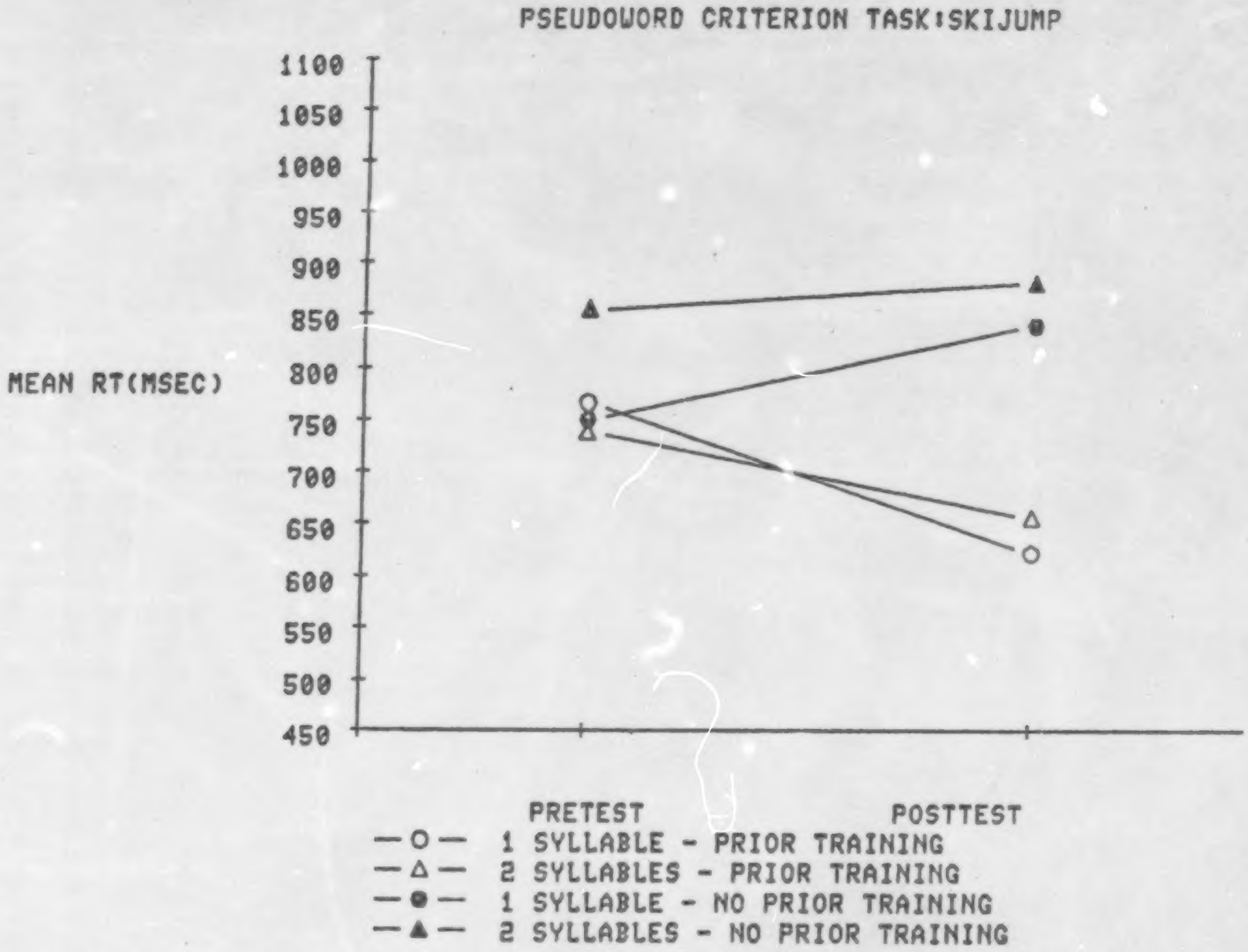


Figure 3.43. Mean pseudoword pronunciation latencies obtained by subjects before and after training on Ski-Jump. Subjects in the prior training group had completed Speed and Racer prior to undertaking Ski-Jump.

the frequency by training by syllable interaction, $F_{1,144}=5.23$, $p<.05$). The other two subjects showed no reductions in RT; one, MG, showed a significant increase in mean RT from 534 to 689 msec ($F_{1,144}=17.1$, $p<.005$). Our tentative conclusion is that although an automated subskill such as decoding may continue to be exercised in the performance of a higher level training task such as Ski-Jump and therefore be further improved, such a training environment is far from an optimal or reliable in developing such subcomponents: direct and immediate feedback correlated with subskill performance is required for efficient training of decoding skills.

Effects of Ski-Jump training on accuracy of pseudoword decoding resulted in a marginally significant 7% improvement in accuracy ($F_{1,5}=4.28$, $p=.09$), and an interaction of frequency and syllable length ($F_{1,5}=5.94$, $p=.06$). Two-syllable pseudowords derived from high frequency words were slightly easier to decode than were the other items. There were no other significant effects of group, nor were there significant interactions of group or training with the other factors. Performance was uniformly accurate with an average of 89%.

There were no significant effects of training on accuracy or speed in the word pronunciation task. In this task, the mean pretest RT for previously untrained subjects was 675 msec, while it was 492 msec for the previously trained group. The posttest latencies for the two groups were 620 msec and 527 msec respectively. The only effect to even approach significance in the analyses of mean RTs was the interaction of group, syllable length, and training ($F_{1,4}$=3.41, p=.14). Two-syllable words produced a greater mean RT (+71 msec) than one-syllable words for two of the previously untrained subjects[**********], while the reverse was the case for the previously trained subjects (-51 msec). There were no significant effects of Ski-Jump training on mean RT or on the magnitude of the syllable effects in the word pronunciation task. In the analyses of word pronunciation latencies for individuals, there were significant reductions in latency following Ski-Jump training for two subjects, one previously untrained, LN (556-470 msec, $F_{1,144}$=11.45, p<.001), and one previously trained, JS (1057-820 msec, $F_{1,144}$=8.68, p<.005).

Span of apprehension. Training on the Ski-Jump system led to a marginally significant improvement in visual span from a mean of 11.5 letter spaces in the pretest to 12.8 letter spaces in the posttest ($F_{1,5}$=3.71, p=.11). In the analyses of mean span for individuals, four subjects showed significant or marginally significant improvements following Ski-Jump training: AC (10.6-14.25 letter spaces, $F_{1,150}$=32.6, p<.001), JS (5.6-8.0 letter

---

[**********] Posttest data for the third untrained subject were lost due to equipment failure.

spaces, $F_{1,150}=12.9$, $p<.005$), MG (13.7-16.4 letter spaces, $F_{1,156}=10.76$, $p<.005$), and LN (14.4-16.0 letter spaces, $F_{1,156}=2.86$, $p<.10$). The other subjects showed no significant changes in visual span; their pre- and posttest measures were 7.9-8.4 (EK), 11.2-11.2 (OD) and 17.2-15.7 (KG). Subjects KG and LN were near the ceiling on the span measure, even on the pretest.

In the analysis of response latencies, there were no significant reductions in mean RT in either the group analysis or the individual analyses. Thus, effects of Ski-Jump training, when they occurred, were on the amount of information encoded within a fixation rather than on the latency in reporting that information. This is in contrast to the effects of the Racer training on this measure. Racer effects were felt more generally on speed of responding, and less generally on width of the visual span.

Inference task. The inference task yielded three criterion measures of reading skill. The first two measures of comprehension were concerned with the ability of a subject to infer a high order relation among sentences that was not explicitly marked and to select a conjunction that appropriately represented that relation. The first measure was the percent of correct selections of an appropriate conjunction and the second was the mean RT for making these selections. The mean percent of correct judgments and mean RTs are given in Figures 3.44 and 3.45, for the two subject groups. Referring to Figure 3.44, there was a significant main effect of training on Ski-Jump ($F_{1,5}=7.81$, $p=.04$) and a significant interaction of group and training ($F_{1,5}=9.67$, $p=.03$). The group that received prior training using Speed and Racer had a higher initial accuracy level, and showed no change in accuracy with practice on Ski-Jump. In contrast, the group not previously trained had a lower initial level of accuracy, and showed a 23% improvement following Ski-Jump training. The results of analyses for individual subjects were consistent with this group result. The previously untrained subjects showed pre- and posttest scores of 50-60% (EK; $Z=.64$, $p=.26$), 55-90% (LN; $Z=2.69$, $p=.003$), and 60-85% (AC; $Z=1.84$, $p=.03$). None of the previously trained subjects showed either significant improvements or declines in accuracy of judging appropriate conjunctions. Mean RTs for the two groups of subjects on this task are shown in Figure 3.45. While there appears to be a substantial difference in means for the two groups, the effect of prior training is not significant ($F_{1,5}=3.81$, $p=.11$). There were small decreases in mean RT for both groups. In the analyses carried out for individuals, two subjects showed significant reductions in mean RT following Ski-Jump training; subject EK improved from 6.37 to 4.35 sec ($t_{38}=3.28$, $p=.001$) and subject MG improved from 2.40 to 2.00 sec ($t_{38}=1.67$, $p=.05$).

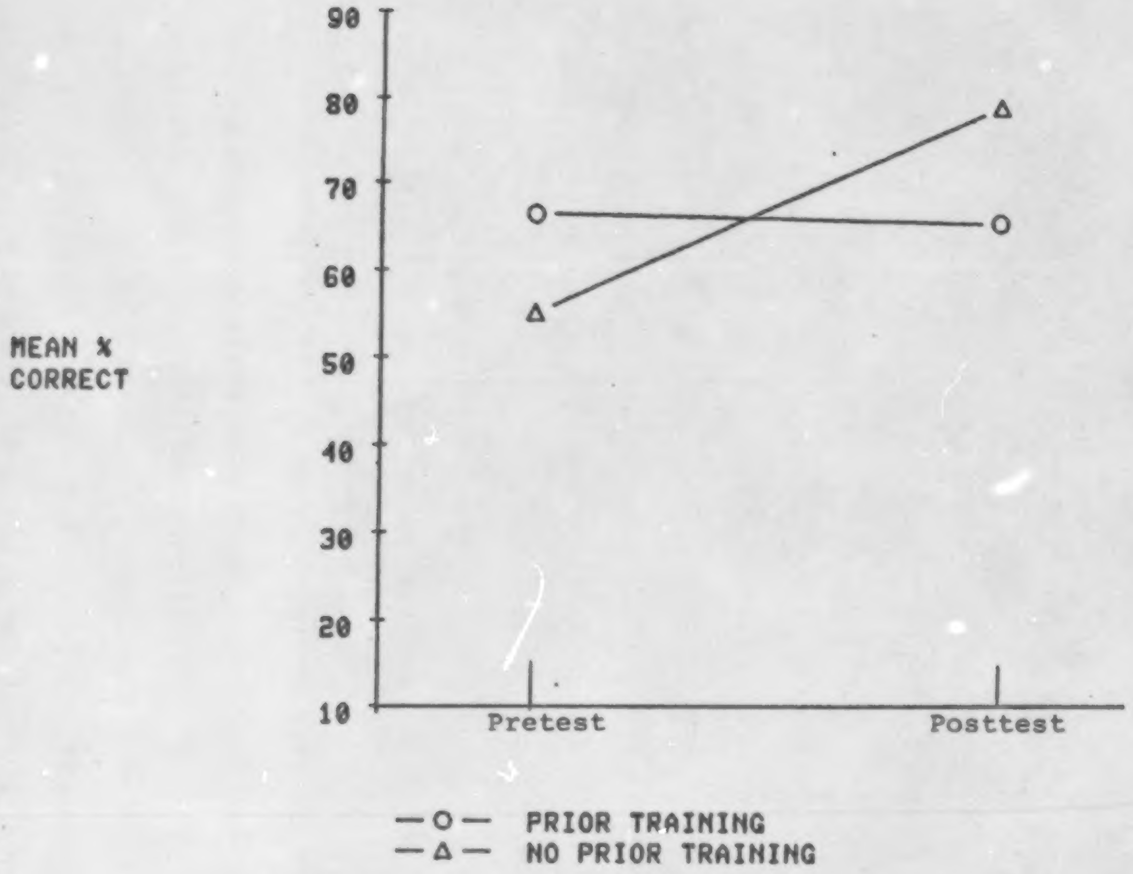


Figure 3.44. Mean percent of correct selections of conjunctive expressions in the inference task obtained by subjects before and after training on Ski-Jump. Subjects in the prior training group had completed Speed and Racer prior to undertaking Ski-Jump.

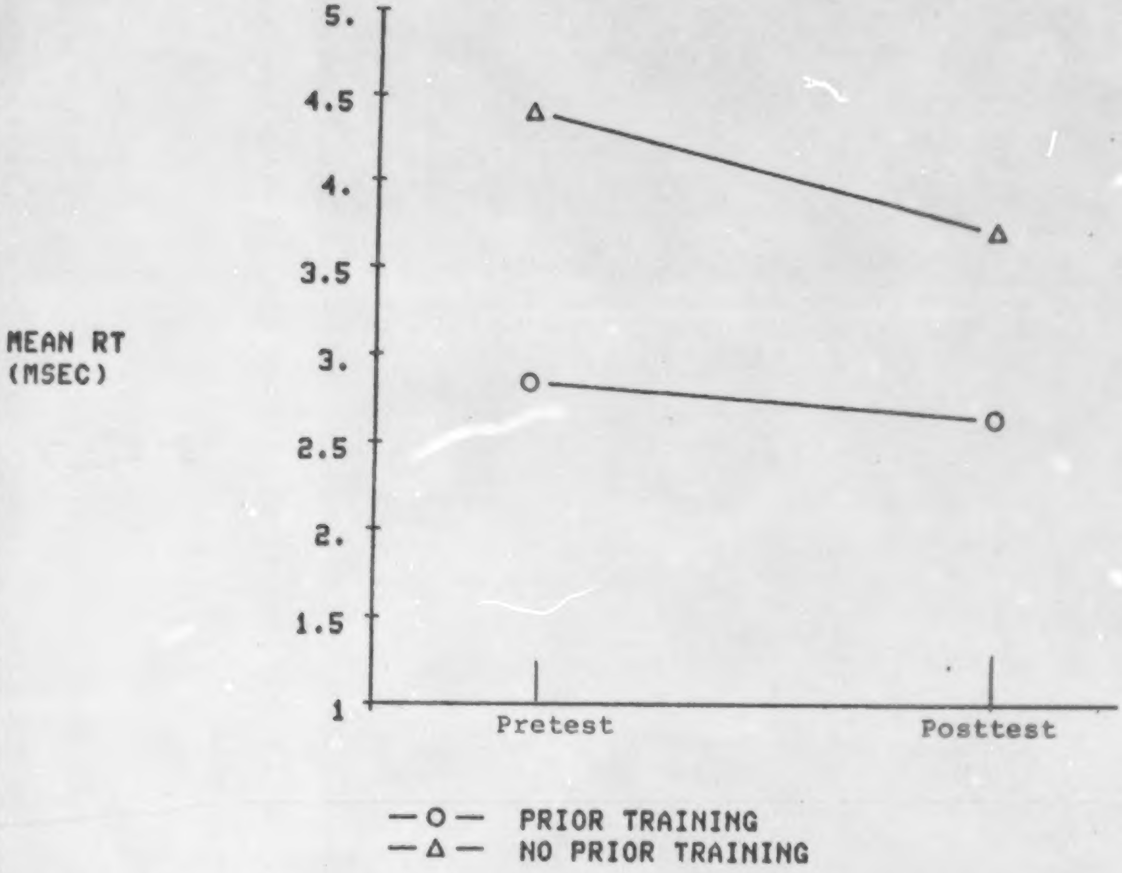


Figure 3.45. Mean RT in judging appropriateness of conjunctive expression in the inference task obtained by subjects before and after training on Ski-Jump. Subjects in the prior training group had completed Speed and Racer prior to undertaking Ski-Jump.

When improvements in performance on the inference task occurred, they were not accompanied by increases in RT for making judgments.

Our final criterion measure was the time required to read the final sentence in the inference task. This measure, in msec per word, gives a general indication of a subject's reading speed (which is actually the inverse of rate) under conditions where there is a demanding comprehension requirement. The results are shown in Figure 3.46. There was a significant main effect of Ski-Jump training ($F_{1,5}=7.26$, $p=.04$) and a significant group by training interaction ($F_{1,5}=7.34$, $p=.04$). Subjects receiving only Ski-Jump training showed no change in reading speed, while subjects who were given the entire sequence of training on Speed, Racer, and Ski-Jump showed an 84% reduction in reading time per word. These same subjects had shown no reduction in reading speed following Racer training. Training in the Ski-Jump task following training in decoding appears to have encouraged consolidation of decoding and comprehension skills, as manifested in increases in reading speed without loss of comprehension in the inference task.

Summary. The Ski-Jump game requires subjects to combine perceptual and word recognition skills, and to simultaneously make use of the semantic context provided by a sentence frame to facilitate lexical identification and semantic interpretation of

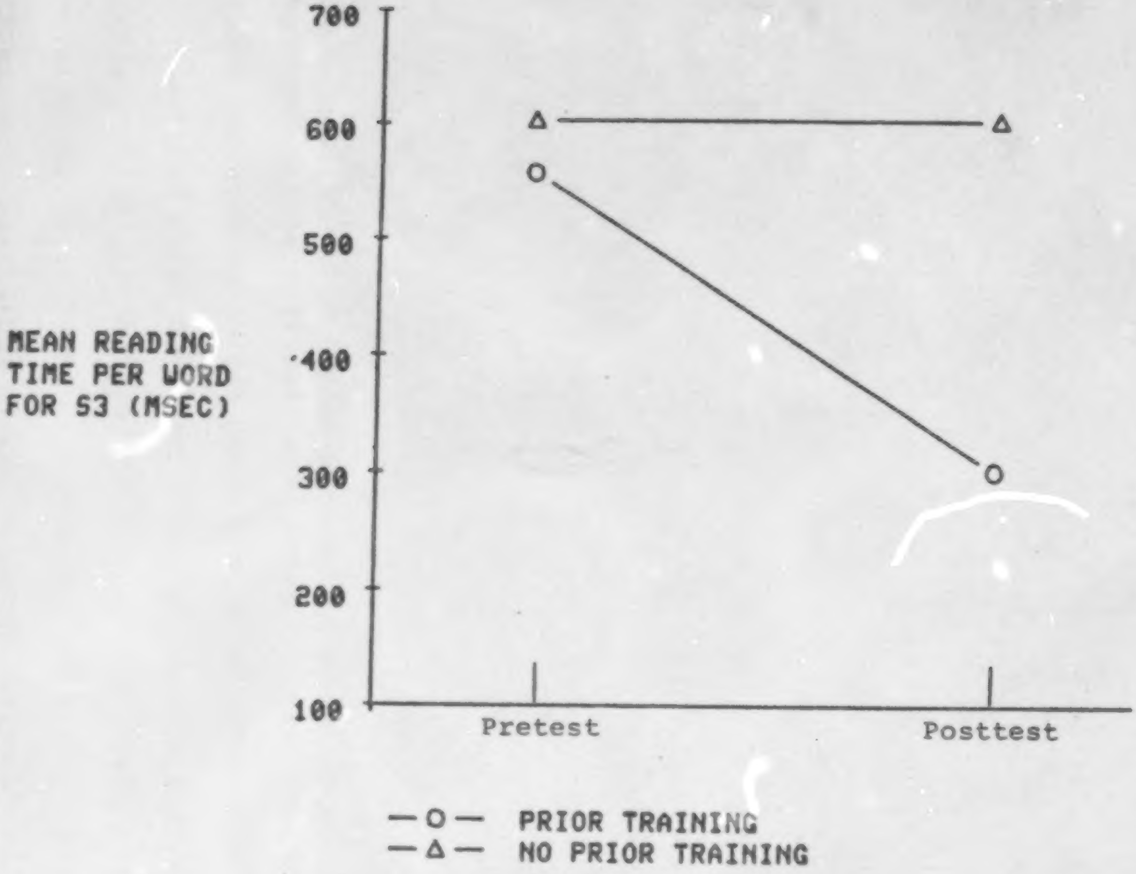


Figure 3.46. Mean reading time in msec per word for the third sentence used in the inference task, for subjects before and after Ski-Jump training. The third sentence contains propositions that are logically, temporally or causally related to proposition in the first two sentences. Subjects in the prior training group had completed Speed and Racer prior to undertaking Ski-Jump.

a target word. Masking of the target word forces the subject to rely on semantic context. As the subject progresses in training, the constraining power of this context is reduced, and the subject must learn to employ more general constraints that focus on semantic classes of words rather than individual lexical categories. Our evaluation of skill acquisition following 10-16 practice sessions using Ski-Jump indicates that all subjects improved in their ability to establish and use frame-based activation of concepts in semantic memory. Ski-Jump training also provided some opportunity to develop perceptual encoding skill, although it was not as effective a training environment as the Speed game for developing that skill. The effects of Ski-Jump training on decoding skill were restricted to subjects who had already had the opportunity to build efficient word decoding skills using the Racer system. Ski-Jump training also had some impact on our more general criterion measures of reading skill. Four of our seven subjects showed improvement in span of apprehension, and four showed improvement in accuracy or RT on the inference task. Since the specific skills required for inferring high order relations among sentences were not addressed by the training systems, we interpret such gains as evidence of forward transfer of lower-level contextual priming skills to higher order analytic processes. For example, priming of concepts in semantic memory might facilitate the tracing of collocative references and the reinstatement of propositions that are linked through such referential relations. These processes could contribute, in turn, to the establishment of high order relations among propositions. Finally, subjects who completed the entire Speed-Racer-Ski Jump training sequence showed significant increases in reading rate, from an average of 108 wpm to 199 wpm with no drop in comprehension (65-66% accuracy), as measured by the inference task. This suggests that improvements in the level of automaticity of multiple skill components can reduce the effort required in reading text for comprehension.

3.5 Cumulative Summary of Performance for Subjects who Completed the Training Sequence.

Cumulative summaries of pretest and posttest performance on each of the criterion measures are given in Tables 3.1-3.4 for each of the subjects who completed the entire training sequence. Subject JS (Table 3.1) showed an improvement in speed and accuracy of unit detection following Speed training, and maintained high levels of performance throughout the remainder of the study. He showed transfer of speed training in his pseudoword pronunciation accuracy, but this gain was not maintained following the 6 month interval between the Speed and Racer training experiments. Racer training resulted in a substantial improvement in accuracy of decoding pseudowords (from 16% to 69%) but with no reduction in pronunciation latency. Subsequent training on Ski-Jump resulted in a substantial

Table 3.1

Cumulative Summary of Performance: Subject JS

Criterion Measure	Speed	Racer		Ski-Jump
	Pretest	Posttest/Pretest	Posttest/Pretest	Posttest
Unit Detection				
RT	323 msec.	288 msec.	246 msec.	253 msec.
ACC	77%	85%	84%	85%
Pseudoword Pron.				
RT	1300 msec.	1467/1522 msec.	1639 msec.	1126 msec.
ACC	13%	41/16%	69%	73%
Word Pronunciation				
RT	--	1119 msec.	1057 msec.	820 msec.
ACC	--	42%	72%	79%
Context Priming				
HCC	--	--	2770 msec.	1000 msec.
LCC	--	--	2267 msec.	1150 msec.
All	--	--	2445 msec.	848 msec.
Span of Apprehension				
No Text	7.9 letters	8.1 letters	5.5 letters	7.7 letters
Text	5.6 letters	9.7 letters	5.8 letters	8.3 letters
Inference				
ACC	--	35%	50%	45%
RT	--	2586 msec.	2071 msec.	2501 msec.
Reading Rate	--	227 wpm	214 wpm	588 wpm
Nelson-Denny				
Total	32	--	--	35
Percentile	10%	--	--	11%

Table 3.2

Cumulative Summary of Performance: Subject MG

Criterion Measure	Speed	Racer	Ski-Jump	
	Pretest	Posttest/Pretest	Posttest/Pretest	Posttest
Unit Detection				
RT	288 msec.	194 msec.	235 msec.	224 msec.
ACC	91%	80%	91%	90%
Pseudoword Pron.				
RT	807 msec.	1022/773 msec.	533 msec.	689 msec.
ACC	62%	78%/66%	75%	95%
Word Pronunciation				
RT	--	697 msec.	428 msec.	465 msec.
ACC	--	78%	89%	98%
Context Priming				
HCC	--	--	543 msec.	382 msec.
LCC	--	--	1060 msec.	727 msec.
All	--	--	847 msec.	546 msec.
Span of Apprehension				
No Text	14.2 letters	15.1 letters	14.2 letters	15.7 letters
Text	15.7 letters	16.7 letters	13.3 letters	16.8 letters
Inference				
ACC	--	80%	85%	75%
RT	--	1961 msec.	2404 msec.	2005 msec.
Reading Rate	--	101 wpm	91 wpm	106 wpm
Nelson-Denny				
Total	39	--	--	61
Percentile	29%	--	--	54%

Table 3.3

Cumulative Summary of Performance: Subject KG

Criterion Measure	Speed Pretest	Speed Posttest/Pretest	Racer Posttest/Pretest	Ski-Jump Posttest
Unit Detection				
RT	362 msec.	194 msec.	217 msec.	256 msec.
ACC	90%	74%	84%	90%
Pseudoword Pron.				
RT	1402 msec.	500 msec.	477 msec.	45[illegible] msec.
ACC	89%	91%	94%	94%
Word Pronunciation				
RT	736 msec.	403 msec.	314 msec.	348 msec.
ACC	90%	98%	98%	98%
Context Priming				
HCC	--	--	1920 msec.	1091 msec.
LCC	--	--	636 msec.	873 msec.
All	--	--	1251 msec.	794 msec.
Span of Apprehension				
No Text	12.9 letters	15.7 letters	17.2 letters	16.2 letters
Text	14.2 letters	14.4 letters	17.2 letters	15.2 letters
Inference				
ACC	40%	65%	55%	65%
RT	2572 msec.	3075 msec.	3555 msec.	3412 msec.
Reading Rate	132 wpm	147 wpm	116 wpm	588 wpm
Nelson-Denny				
Total	30	--	--	22
Percentile	14%	--	--	6%

Table 3.4

Cumulative Summary of Performance: Subject OD

Criterion Measure	Speed	Racer	Ski-Jump	
	Pretest	Posttest/Pretest	Posttest/Pretest	Pretest
Unit Detection				
RT	315 msec.	182 msec.	174 msec.	171 msec.
ACC	91%	85%	83%	78%
Pseudoword Pron.				
RT	595 msec.	398 msec.	355 msec.	289 msec.
ACC	80%	95%	98%	99%
Word Pronunciation				
RT	589 msec.	306 msec.	341 msec.	312 msec.
ACC	85%	96%	94%	98%
Context Priming				
HCC	--	--	3278 msec.	1524 msec.
LCC	--	--	1917 msec.	718 msec.
All	--	--	2597 msec.	1121 msec.
Span of Apprehension				
No Text	8.8 letters	9.5 letters	10.4 letters	11.3 letters
Text	9.5 letters	8.3 letters	12.0 letters	11.1 letters
Inference				
Acc	70%	65%	75%	75%
RT	6132 msec.	5136 msec.	3315 msec.	2611 msec.
Reading Rate	89 wpm	79 wpm	78 wpm	137 wpm
Nelson-Denny				
Total	36	--	--	--
Percentile	26%	--	--	--

decrease in decoding latency with yet another small increase in accuracy (to 73%). JS's accuracy in word pronunciation increased from 42% to 72% following Racer training and his latency in pronunciation showed a decrease as a result of Ski-Jump training. Ski-Jump training also resulted in a substantial reduction in latency in the context priming task on the critical (low probability) target words. This reduction occurred for both high and low constraining contexts. On the span of apprehension task there was an overall increase in span for the context condition but the gain was modest in size. Subject JS showed no changes in accuracy or latency in the inference task, and no change in Nelson-Denny test scores (total of comprehension and vocabulary). Finally, though he showed an increase in reading rate, it occurred for a task in which his measured comprehension (performance in judging appropriate connectives in the inference task) was low. In summary, this subject has shown substantial improvement in the perceptual, decoding, word recognition, and context utilization components of reading; however these gains have not been reflected in his comprehension and vocabulary test scores.

Subject MG (Table 3.2) showed a reduction in RT for unit detection following Speed training, accompanied by an increase in accuracy and latency for pseudoword decoding. Racer training resulted in further improvement in accuracy and a dramatic reduction in latency for decoding words and pseudowords; however, these improvements were accompanied by an increase in unit detection latency. Following training with Ski-Jump, word and pseudoword decoding latencies remained low and accuracy of decoding increased to 95%-98%. Ski-Jump training also brought a substantial reduction in RT for recognizing and judging the semantic acceptability of target words in the context task and an increase in visual span. No substantial changes in performance on the inference task were shown, although this subject showed a marked improvement in Nelson-Denny total score, from 39 to 61. His final score corresponds to the 54th percentile.

Subject KG, although falling within the 14th percentile on the Nelson-Denny test, showed high levels of accuracy in decoding words and pseudowords as well as a wide span of apprehension in initial pretests. However, her initial level of comprehension on the inference task was low. Speed training resulted in a substantial reduction in unit detection latencies, in part due to an acceptance of more errors (her accuracy dropped from 90% to 74%). Her subsequent performance in the unit detection task showed a small increase in RT (60 msec) but a return to the initial high level of accuracy (90%). For KG, Speed training resulted in a dramatic reduction in latency for decoding both words and pseudowords, accompanied by increases in accuracy. Racer training resulted in a further decrease in latency which was maintained following Ski-Jump training. Her final levels of performance in the decoding tasks were extremely high. Speed training also resulted in an improvement in her span of apprehension under the no context condition, and in an inrease in accuracy from 40% to 65% on the difficult inference test. These performance levels were again achieved at the end of training. Ski-Jump training primarily affected her performance on the context priming task and her reading rate, which increased dramatically. However, there was a small decline in her score on the Nelson-Denny test.

Our last subject, OD (Table 3.4), initially showed a high level of accuracy in decoding (80-85%), but had a narrow span of apprehension and a slow reading rate in the inference task. Training with the Speed system brought a substantial improvement in RT in unit detection; this level of performance was maintained over the 6 month span of the study. Speed training also resulted in a decrease in pseudoword and word decoding latencies, accompanied by substantial improvements in accuracy. Decoding accuracy continued at these high levels (94-98%) following Racer training, and following Racer and Ski-Jump practice there were additional reductions in RT for decoding pseudowords. Ski-Jump training also brought about a dramatic improvement in time for recognizing and integrating target words in the context task and an increase in reading rate. There were reductions in time for making semantic decisions about the appropriateness of connectives in the inference task following training on each of the games. However, the proportion of correctly judged connectives did not change substantially.

4. DISCUSSION

4.1 Nature of the Skills Acquired

The design of our training study and, in particular, the use of a battery of criterion tasks enabled us to determine with some specificity the nature of the skills acquired in each training system.

The Speed game was designed so that successful performance necessitated the development of perceptual skills. These could be of several types:

1. Unit detectors. Specific unit detectors may be developed for each multiletter unit trained. These detectors are postulated to have properties similar to cognitive demons in the Pandemonium System (Selfridge, 1959) which act as mini-productions, performing their function whenever defining input states are encountered (e.g., visual features). According to this view, improvements in performance should be specific to the units trained, but not be limited to the detection task.

2. Strategic application of prior orthographic knowledge. Trainees may be learning to apply strategically their prior knowledge of unit positional likelihoods to focus their attention on specific portions of the target words in the detection task. Such a strategic skill will suffice for those conditions in which a unit, trained or untrained, appears reliably in the beginning or end of a word, but will not allow for successful performance for units whose positions are unpredictable. This strategy should result in successful performance on a unit detection task, but should not transfer to tasks in which units are not specified in advance.

3. Shifts in criterion for detection. Trainees may develop more lenient decision criteria for detecting

target units. A shift in decision criteria can result in performance improvements in detecting units that have not previously been trained as well as for trained units, but the improvements should be reflected in decreased RT accompanied by increases in the number of false unit detections, that is, in the number of errors. Shifts in detection criteria may benefit the performance of tasks other than detection tasks if, in general, the criteria of unit detectors have been modified as a result of training. For subjects who start out with high levels of accuracy, this criterion shift might result in an improvement in efficiency of unit recognition, at the expense of only small increases in rates of false detections.

4. Allocation of attention. Trainees may be learning to allocate attentional resources to improve efficiency of perceptual encoding, and to distribute their attention across letter positions within a target word so as to more rapidly detect units in the more difficult medial positions. Both would serve to improve the quantity and quality of perceptual information that is encoded under perceptually demanding conditions. Such a skill would not be limited to the set of units that have been trained, or to the detection task. Thus, it could contribute to successful performance in any task requiring rapid encoding of orthographic information in a visual array.

The results for our training task and, more particularly, the results for our transfer tasks help us to decide among these alternatives.

If it were the case that students were developing only unit detectors, we would expect their performance on trained units in the detection task to be superior to that for untrained units. However, our results clearly show comparable improvements in performance for the trained and untrained units. Since the gains on the trained and untrained units are of similar magnitude (132 and 91 msec, respectively), we conclude that an explanation based solely on the development of unit detectors is unsatisfactory.

The second possibility is that subjects are learning a strategy for applying their knowledge of orthographic groupings. This leads to the prediction that successful performance on the unit detection task will be limited to units occurring in predictable positions. However, the results rule against this possibility. Performance gains as a result of training were as large for the difficult units appearing in medial positions within words as they were for more regular units appearing in initial positions. The average RT gain for trained two-letter units in medial positions was 133 msec and for units in initial positions it was 131 msec. Thus, performance gains for difficult units in less predictable positions were equal to those for easy units in predictable positions. In addition, if students have acquired a strategy based on applying orthographic knowledge, then their performance gains as a result of training should be limited to detection tasks, in which a target unit is specified for each trial. However, there was evidence of transfer to criterion tasks involving pseudoword decoding and word recognition and to the span of apprehension task, in which four of the five subjects showed gains in either the amount of information reported or RT. In none of these transfer tasks were target orthographic units pre-specified.

There was evidence that subjects were developing more lenient detection criteria for units. In general, while there were no significant changes in frequency of correct detections of targets following practice, there was a significant drop in accuracy for similar foils, indicating a more lenient criterion for detecting units. The lowering of criterion was not so large as to increase error rates for dissimilar foils.

Finally, the results suggest that skill acquired in training with the Speed system is based upon a change in the students' allocation of attention. The demonstrated ability of trainees to rapidly detect units when they are embedded within target words, with a mean RT (266 msec) approaching that for units appearing at the beginning of words (238 msec), provides direct evidence of a change in the distribution of attention. Results were even more dramatic for one subject, JS, who was unable to detect units in medial positions of words prior to training. In the pretest, his performance was essentially at the chance level (59% correct). Training brought an improvement in his ability to detect units embedded within an array, to an accuracy level of 80%. Finally, the similarity in the unit detection performance for trained and untrained units, along with the generalization of training effects to criterion tasks that do not explicitly involve unit detection, support the view that the attentional skills acquired are more general than the development of a specific unit vocabulary and are applicable in tasks other than the detection task.

The Racer game requires subjects to respond orally with the correct pronunciations of stimulus words. However, there is no direct confirmation of the correctness of either the pronunciation or the lexical identification of the stimulus. The intent was to build a practice environment in which the focus was on efficient phonological decoding of stimuli, with the Sound Trap ensuring that a reasonable effort would be made at correct decoding. Errors were also minimized by adopting a training strategy of having the subject work from easy or consistent sets of words to more difficult, mixed matrices of words. While the focus was on establishing efficient decoding skills, there are a number of skill components of reading, as in the case of the Speed system, that could be developed as a result of Racer training:

1. "Pure" decoding. Subjects may develop automated procedures for applying rules for translating orthographic information into a phonological or articulatory representation for the stimulus. If such a skill is developed, performance gains due to Racer training should be apparent for a pseudoword pronunciation task as well as for a word pronunciation task, despite the fact that Racer training was restricted to words. Moreover, the gains shown in pseudoword pronunciation should be similar in magnitude to those for words, and should be reflected in reductions in the length and syllable effects associated with words and pseudowords.

2. Rapid word recognition and "decoding by analogy". Subjects may have acquired an ability to rapidly identify words on the basis of their visual and orthographic features, and to efficiently retrieve the phonological and articulatory information available for the word in memory. If a stimulus word is unfamiliar, activation of familiar words that are orthographically and phonologically similar to the stimulus will facilitate the retrieval of phonological and

articulatory information relevant to the pronunciation of the unfamiliar stimulus (cf. Glushko, 1980). This information in turn could provide a basis for building a phonological representation for the unfamiliar stimulus, through a process of "debugging" or modifying the pronunciation of the analogous word. If it is skill in word recognition that is developed in Racer training, then the gains in performance for words should be greater than those for pseudowords, as the training task employs only real words as stimuli, and words are all drawn from a vocabulary of moderate to high frequency items that are not likely to be unfamiliar to subjects. Rapid activation of word units on the basis of visual information should also lead to more efficient encoding of orthographic information in a word (Rumelhart and McClelland, 1981), and should thus lead to improvement in performance on the unit detection task and on the span of apprehension task.

3. Pronunciation strategies. Subjects may, as a result of Racer training, develop strategies for initiating their vocalization prior to completion of decoding processes. Such a strategy is limited to some extent by phonotactic co-occurrence rules. Development of an early pronunciation strategy would lead to greater effects of Racer training on RTs for longer stimulus items than for shorter (1-syllable) items. Furthermore, there would be no effect of training on accuracy of decoding. And, a pronunciation strategy, while leading to gains in speed on the Racer task, should not lead to improvement on the criterion word and pseudoword pronunciation tasks, since in those tasks RTs were measured for one item at a time under non-speeded conditions and the accuracy of pronunciation was monitored by an experimenter. Finally, there should be no general transfer of training to tasks in which vocal responses are not required.

The results of our training experiment allow us to assess which of these skills are developed as a result of practice using the Racer system. There is strong evidence for the development of "pure" decoding skills. The reductions in latency for pronouncing pseudowords were as great as or greater than those for words. There were increases in accuracy of pronunciation for both words and pseudowords. And, the effects of syllable length on performance in the pronunciation tasks were smaller after training. None of these results are compatible with the view that the skills acquired are solely those of rapid word recognition. Performance improvements for words were not greater than those for pseudowords: there was little transfer to the unit identification task, and there was no increase in span of apprehension following Racer training, although RTs in this task were reduced. Finally, while it is clear that the development of an early pronunciation strategy was employed in the Racer task and that it enabled subjects to reach extremely high speeds during training, the transfer of skills acquired in Racer training to the criterion measures of decoding for words and pseudowords suggests that more general decoding abilities were developed as well.

The Ski-Jump game requires subjects to utilize both the constraining semantic information contained in a context and the perceptual information contained in the masked stimulus words to make lexical identifications. It also requires subjects to evaluate the meaning of lexical items in terms of their appropriateness in a given context. The training task may thus serve to foster the acquisition of more than one skill component:

1. _Pure perception_. Subjects may develop an ability to encode orthographic units with minimal visual evidence, using the sequential redundancies of letter groups in

words to achieve higher accuracy and efficiency of encoding under perceptually demanding conditions. If this skill is the one developed, subjects should show improvement in accuracy and latency on the unit detection task as a result of Ski-Jump training.

2. Efficient word recognition. Subjects may develop efficient and accurate skills for recognizing words on the basis of their visual features. Such a skill should be manifested in an improvement in word recognition efficiency, and in an increase in span of apprehension as a result of Ski-Jump training. Skill improvements on the context priming task should be similar for high and low probability words and for foils (unrelated words).

3. Frame-based activation of concepts. Subjects may acquire an ability to make use of the semantic information provided by a sentence frame to activate relevant concept nodes in semantic memory and the lexical items attached to those nodes. Subjects may also develop an ability to utilize such activations to facilitate lexical identification of target words. If such skills are acquired in Ski-Jump training, subjects should show greater improvement for relevant target words than for foils. If the contextual priming process involves a truly "parallel" activation of the multiple lexical categories associated with a context-relevant concept, then the final level of performance for low probability stimulus words should equal that for high probability items. Skill in using sentence frames to activate concepts in semantic memory constitutes a necessary, but not sufficient condition for improvement on tasks involving the tracing of referential relations. Since tracing of referential relations may in turn lead to reinstatement of propositions containing the referred items, an improvement in ability to analyze relations among propositions might also result.

4. Analysis of sentence frames (comprehension.) Lastly, subjects given Ski-Jump training may develop an ability to analyze the propositions contained within a context frame, and exploit the propositional relations within the frame to develop a set of semantic constraints that can be utilized in making lexical identifications. While no specific instruction in sentence comprehension is offered in the Ski-Jump game, the game provides a situation in which improvement in such skills would facilitate game performance. Such skills, if they are

developed, should lead to improved accuracy in judging appropriate connectives in the inference task.

The results of our training experiment again allow us to determine the skill components that are showing the most development as a result of practice. The most striking result of Ski-Jump training was the improvement in performance on the context priming task that occurred for both high and low probability targets. In addition, a clear feature of the training records obtained during Ski-Jump practice was the difference in effects of practice on target words and unrelated words. Improvements were far more modest in the latter category than in the former one. Taken together, these results support the interpretation that subjects have developed an ability to employ the semantic information contained within sentence frames to activate frame-related concepts in memory, and to use such activations as an aide in recognizing and interpreting stimulus words that are presented under visually degraded conditions.

We also found improvement in performance on the unit detection task for trainees who had not reached asymptotic levels of performance on the task prior to Ski-Jump training. This result is consistent with the possibility that perceptual and/or word recognition skills are developed. However, the lack of improvement for those subjects on the word pronunciation and the pseudoword decoding tasks suggests that of these it is the perceptual component that is the one actually developed.

Finally, there was some evidence of improvement on the inference task as a result of Ski-Jump training. This is consistent with the interpretation that subjects are improving in their ability to analyze sentence frames similar to those used in the Ski-Jump task. However, the modest performance gains on the inference task are also consistent with the development of skill in frame-based activation of concepts, and as there is other independent evidence for the development of such a skill, it is more parsimonious to attribute improvement on the inference task to the context utilization component.

4.2 Principles for Optimizing Transfer of Training in a Skill Hierarchy

The patterns of transfer of skills acquired in the Speed, Racer, and Ski-Jump systems observed for groups of trainees who differed in their prior training histories lead to some conclusions regarding optimal training environments and optimal sequencing of training exercises, including the performance criteria that should be used in determining mastery at each stage of training.

Optimal training environments. Our group of trainees consisted of students who have not acquired adequate reading skills over approximately ten years of schooling in which they received conventional reading instruction. For such subjects, our initial question concerned the feasibility of constructing training environments in which individual skill components of reading could be developed with a modest expenditure of time, typically amounting to 2 hours a week for a period of 4 to 6 weeks on a single training system. Dramatic improvements in the targeted skills were reflected not only in performance on the training games themselves, but in tests for transfer of training as well. With carefully designed training environments it was possible for our subjects to reach performance levels that equal or exceed those of high ability readers who have not had the benefits of training.

Two characteristics of our games account, in large part, for their effectiveness in training skilled performance: the focus on developing particular skill components, and the delivery of feedback that is both immediate and directly correlated with the skilled performance. In addition, the game-like character of the training systems motivated subjects to engage in practice that would otherwise be dull and routine.

Strong support for the importance of providing immediate feedback specifically correlated with the skilled performance is found in the fact that improvements in skills specifically targeted by a training game were far more dramatic than improvements in those that were not so targeted. For example, Ski-Jump training was much less effective in developing the perceptual skills measured in the unit detection task than the

Speed system; Ski-Jump was also less effective in developing efficient decoding skills than the Racer system. In fact, only those subjects who had received prior training in decoding showed any improvement in decoding skill as a result of Ski-Jump training.

Transfer of training: Higher to lower order components. A low level component, when trained to automatic but not asymptotic levels, will be reliably executed in the performance of training tasks focused on higher level components whenever the higher level component is functionally linked to the lower level component. In such a situation, further improvement on the low level component can take place while training focused on the higher level component proceeds. And, the demands on the lower skill made by the training task will not interfere with acquisition of the targeted higher level skill. Evidence for the development of a lower level component in the context of training a higher level component was found in our evaluation of the Ski-Jump game. We found reductions in pseudoword pronunciation latencies for a number of our subjects who had received prior Racer training. These subjects apparently employed their automatic decoding skills while performing the Ski-Jump task (which was a silent reading task), and the exercise of those decoding skills in the new context allowed them to improve further in their decoding abilities.

We did not find similar instances of skill improvement in the unit detection task for subjects who had received prior Speed training as they completed training on the Racer and Ski-Jump games. We believe that these subjects had already reached what are essentially asymptotic levels of performance on the perceptual skill following Speed training, leaving little room for improvement in that skill in their subsequent Racer and Ski-Jump training. However, we may conjecture that had these subjects terminated their Speed training before reaching asymptotic levels of performance, further improvements in the perceptual skill would have occurred during their subsequent training activities.

Transfer of training: Lower to higher level components. Training of low level components to automatic levels will have an impact on the performance of higher level skills that are functionally related to the trained component. These skill interactions follow the hypothesized patterns of skill interaction developed in our earlier analysis of covariances among skill measures (Frederiksen, 1982). For example, Speed training had an influence on accuracy and, for some subjects, speed of decoding. Racer training had an influence on pseudoword as well as word pronunciation efficiency, and on latencies in the span of apprehension task. The nature of the transfer effects appear to depend upon the processing "bottleneck" for the individual subject. For example, subjects who were initially extremely inaccurate in detecting units embedded within words also had high error rates in pronouncing pseudowords. For these subjects, training of the perceptual encoding skill led to increased accuracy in decoding pseudowords, but with increases in pronunciation latencies. It was only with subsequent training using the Racer program that these latencies were reduced. Other subjects, who were initially accurate in encoding orthographic information, showed decreases in latencies for pronouncing pseudowords following training in decoding. Thus, the former subjects were limited in their decoding accuracy by the amount of orthographic information they could encode, while the latter group was constrained by the rate at which such information became available. For both groups, however, there was evidence of skill transfer.

Transfer of training to performance of composite tasks. Training for automaticity of components can eliminate resource bottlenecks in the performance of composite tasks that involve those components. The degree of improvement in composite task performance will depend upon the number of subskills that are performed automatically. Our clearest evidence of the cumulative effects of subskill training on performance of a composite task was seen in our measure of reading rate. Subjects having only Speed and Racer training, or who were trained with Ski-Jump alone, showed no increases in reading rate. Improvement in reading rate occurred only for subjects who completed the entire training sequence. This cumulative effect of training could be due to the development of the three individual skills, or to the integration of those skills as the lower level components are performed during the training of higher level components.

Establishment of skill hierarchies. The order in which training on the three systems was conducted was not arbitrary. Our decisions concerning sequencing of the training exercises were based upon an analysis of interactions among skill components and the establishment of a skill hierarchy. Skill components low in the hierarchy are those that furnish the conditions for improved efficiency and/or accuracy of performance of higher level skills. Thus perceptual encoding precedes decoding as a focus of training, and development of automatic decoding precedes training in the use of context. We believe that further extrapolations of this hierarchical analysis can be made within the domain of skills involved in comprehending discourse. Lower level skills are trainable components that have an influence on the efficient and accurate performance of higher level processes in discourse understanding. These functional interactions are likely to take the form of necessary conditions for higher order processes, and to involve skills that contribute to the establishment of those conditions. For example, following Kintsch and Van Dijk (1978), we postulate that reinstatement of antecedent propositions into working memory is a necessary condition for the analysis of relations among propositions that do not occur together within a text. One means for reinstating propositions is the tracing of referential relations--argument repetitions, lexical references, or other forms of anaphora (cf. Halliday & Hasan, 1976). One skill component that may contribute to efficient and accurate reference tracing is the utilization of frame-based semantic information to prime concepts in semantic memory. Training focused on reference tracing thus constitutes a prime candidate for the future extension of the reading skills training system.

REFERENCES

Frederiksen, J.R. Sources of process interaction in reading. In A.M. Lesgold and C.A. Perfetti (Eds.), Interactive Processes in Reading. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1981.

Frederiksen, J.R. A componential theory of reading skills and their interactions. In R.J. Sternberg (Ed.), Advances in the Psychology of Human Intelligence, (Vol. 1). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1982.

Frederiksen, J.R., Warren, B.M. and Weaver, P.A. Reader differences in contextual priming of semantic categories. In preparation, 1983.

Frederiksen, J.R., Weaver, P.A. and Warren, B.M. Text-based inferential processing in readers of varying ability. In preparation, 1983.

Gillotte, H.P. Training in the generative use of context for improving integrative processing skills. Unpublished doctoral dissertation, Harvard University, 1983.

Glushko, R.J. Principles for pronouncing print: The psychology of phonography. In A.M. Lesgold and C.A. Perfetti (Eds.), *Interactive Processes in Reading*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1981.

Halliday, M.A.K. and Hasan, R. *Cohesion in English*. London & Longman, 1976.

Harris, A.J.and Sipay, E.R. *How to Increase Reading Ability*. New York: McKay, 1975.

Kintsch, W. and vanDijk, T.A. Towards a model of text comprehension and production. *Psychological Review*, 1978, *87*, 329-354.

Malone, T.W. Toward a theory of intrinsically motivating instruction. *Cognitive Science*, 1981, *5*, 333-369.

Mayzner, M.S. and Tresselt, M.E. Tables of single-letter and digram frequency counts for various word-length and letter position combinations. *Psychonomic Monograph Supplements*, 1965, *1*, 13-32.

Perfetti, C.A. and Lesgold, A.M. Discourse comprehension and sources of individual differences. In M.A. Just and P.A. Carpenter (Eds.), *Cognitive Processes in Comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Perfetti, C.A. and Roth, S. Some of the interactive processes in reading and their role in reading skill. In A.M. Lesgold and C.A. Perfetti (Eds.), Interactive Processes in Reading, Hillsdale, N.J.: Lawrence Erlbaum Associates, 1981.

Rumelhart, D.E. and McClelland, J.L. Interactive processing through spreading activations. In A.M. Lesgold and C.A. Perfetti (Eds.), Interactive Processes in Reading. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1981.

Schneider, W. and Shiffrin, R.M. Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review, 1977, 84, 1-66.

Selfridge, O.G. The Mechanization of Thought Processes. London: H.M. Stationary Office, 1959.

Shiffrin, R.M. and Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 1977, 84, 127-190.

State of New York, The Board of Regents. Degrees of Reading Power Test. Albany, New York: State Education Department, 1977.

Venesky, R.L. *The Structure of English Orthography*. The Hague: Mouton, 1970.